DOCTOR OF PHILOSOPHY

Targeting an E3 ubiquitin ligase Siah1 and a cysteine protease SENP1 using SPR and DSF-based fragment screening

Rimsa, Vadim

Award date:
2013

DOCTOR OF PHILOSOPHY

Targeting an E3 ubiquitin ligase Siah1 and a cysteine protease SENP1 using SPR and DSF-based fragment screening

Vadim Rimsa

2013

University of Dundee

A thesis submitted for the degree of Doctor of Philosophy

Targeting an E3 ubiquitin ligase Siah1 and a cysteine protease SENP1 using SPR and DSF-based fragment screening

Vadim Rimsa

Supervisors:

Prof. William N. Hunter (University of Dundee)

Dr. Derek Ogg (AstraZeneca)

The Wellcome Trust Biocentre

University of Dundee

Scotland

November 2013

Contents

List of figures

List of tables

Acknowledgements

I would like to cordially thank my supervisors Prof. William N. Hunter at University of Dundee and Dr. Derek Ogg at AstraZeneca for the opportunity to carry out my PhD and for their continuous support and expert advice throughout it.

I would also like to thank all the members of the WNH group and the Biophysics group at Alderley Park (AstraZeneca) for making my time enjoyable in both places and for general day to day assistance in the labs. In particular, I am grateful to Thomas Eadsforth for his insightful advice and support on many occasions. I thank Sharon Shepherd for help with protein expression and purification, Paul Fyfe for assistance with X-ray data collection and structure solution. Special thank you is also to Navratna Vajpai, Kevin Embrey and Gareth Davies for useful discussions and help with NMR and SPR experiments. I would also like to acknowledge the BBSRC and AstraZeneca for funding my research.

Finally, I received invaluable support and encouragement during my studies from my family and friends and I would like to thank them immensely, especially my mother Liudmila.

Declaration

I hereby declare that the research presented in this thesis is my own work and it has been written by me. Work carried out by others is acknowledged in the text by reference to the appropriate researchers and publications. No portion of the work presented in this thesis has been submitted in support of any other qualification at this or any other higher education institution.

Vadim Rimsa

The work presented in this thesis is the original work of V. Rimsa. I certify that the candidate has fulfilled the conditions of the relevant Ordnance and Regulations.

Prof. W.N. Hunter

Summary

This dissertation presents fragment screening studies against two human proteins Siah1 and SENP1, which function in post-translational modification pathways. Siah1 is an E3 ubiquitin ligase that functions as a scaffold to transfer ubiquitin bound to an E2 ubiquitin-conjugating enzyme to a substrate as part of the ubiquitination pathway. SENP1 is a cysteine protease that catalyses two essential reactions in the SUMO pathway. It processes pre-SUMO proteins to their mature form and removes SUMO from the target proteins. Siah1 interactions with other proteins involve large surface areas, while SENP1 has a small active site, making it hard to identify ligands for these proteins. The fragment-based approach has emerged as a complementary method to high-throughput screening of finding novel small molecules. The main aim of the study was to examine whether fragment screening would identify any ligands against these targets.

Chapter 1 introduces post-translational modifications and presents fragment-based approach used in drug discovery. Chapter 2 describes the experimental methods used. The results from fragment screening against Siah1 using SPR and DSF are reported in chapter 3. The chapter also presents the structure of Siah1 refined to 1.95 Å that displays new parts of the structure, previously missing due to the absence of reliable electron density. Chapter 4 contains results from the fragment screens against SENP1 using DSF and NMR. The crystal structure of SENP1 was determined with a number of improvements made over earlier structures.

Besides performing fragment screening, the binding between Siah1 interacting proteins reported in the literature and Siah1 was investigated. A number of Siah1 binding partners were successfully expressed and purified as described in chapter 5. One of

those, SIP showed a clear interaction with Siah1, as observed by the shift on a size exclusion column of the complex relative to the individual protein species. Siah1 was reported to collaborate with PEG3 in the regulation of β-catenin degradation. A SCAN domain, located at the N-terminus of PEG3, was tested for binding using gel filtration chromatography and NMR, but no interaction was observed. PEG3 was used in the crystallographic studies and a structure of its SCAN domain was solved using molecular replacement and refined to 1.95 Å. The structure of PEG3-SCAN domain revealed a stable homodimer with an extensive dimerization interface.

The structure of a zinc-dependent cytosolic carboxypeptidase from *Burkholderia cenocepacia* was determined and is reported in chapter 6. This work was a side project assessing a new refinement strategy, which involved the use of the automated protocols embedded in the PDB_REDO server. The structure revealed that carboxypeptidase is a tetramer and provides details of its active site, whose spatial conformation of residues supports the notion that the protein might function as a deglutamylase.

Abbreviations

AZ	AstraZeneca
*Bc*CCP	Cytosolic carboxypeptidase from *Burkholderia cenocepacia*
CCP4	Collaborative computational project number 4
DALI	Distance matrix ALIgnment
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
ESRF	European synchrotron radiation facility
IPTG	Isopropyl β-D-thiogalactoside
ITC	Isothermal titration calorimetry
kbp	Kilo base pair
kDa	Kilodalton
K_d	Dissociation constant
LB	Lysogeny broth / Luria-Bertani
MALDI-TOF	Matrix-assisted, laser desorption ionisation time-of-flight
MES	2-(*N*-morpholino)ethanesulfonic acid
MPD	2-methyl-2,4-pentanediol
MW	Molecular weight
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
OD	Optical density

PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEG3	Paternally expressed gene 3
Pfu	DNA proof reading polymerase from *Pyrococcus furiosus*
PISA	Protein Interfaces Surfaces and Assemblies
RING	Really interesting new gene
RMSD	Root mean square deviation
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
MALS	Multi-angle light scattering
SENP	Sentrin-specific protease
Siah	Seven-in-absentia homolog
SPR	Surface plasmon resonance
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TLS	Translation/libration/screw
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane

Chapter 1

Introduction: Targeting post-translational modifications using fragment-based approach

1.1 Preface

Most proteins undergo some form of chemical alteration following translation. These post-translational modifications confer different properties to the proteins, thereby generating protein heterogeneity. Two such modifications are ubiquitination and SUMOylation, which regulate nearly every aspect of eukaryotic cell biology, including gene transcription, cellular localisation and protein stability. These modifications are carried out by a set of different proteins. Two specific proteins called Siah1 and SENP1; belonging to an E3 ubiquitin ligase and cysteine protease families are introduced in the following chapter. Accumulating evidence suggests that their misregulation is associated with the development of various cancers. Specific potent inhibitors of Siah1 and SENP1 are therefore required to validate these as therapeutic targets.

Identification of enzyme inhibitors and protein ligands to modulate biological function is the foundation of drug discovery. One approach to identify the first ligands is to screen large libraries of compounds with molecular weights of around 500 Da, known as high-throughput screening (HTS). However, practice showed that hits identified are difficult to optimise further. An alternative approach involves the screening of compounds of low complexity, known as fragment-based method. This differs from HTS in terms of library size and screening methods, and hits identified provide better starting points for subsequent drug development. Fragment-based methodology has been successful against some targets like kinases and is now expanding to other protein classes. Herein the fragment screening process is described and the question is asked whether this method can generate ligands targeting protein-protein interactions and poorly accessible active sites such as present in Siah1 and SENP1 proteins.

1.2 Post-translational modifications

Post-translational modification (PTM) is the covalent attachment of a chemical group or a peptide to the protein following translation (Walsh *et al*., 2005). PTMs increase the functional diversity of the proteome and provide an additional level of protein control, which is often reversible. The common modifications include phosphorylation (Ubersax and Ferrell, 2007), glycosylation (Spiro, 2002), S-nitrosylation (Gaston *et al*., 2003), methylation (Grewal and Rice, 2004) and N-acetylation (Glozak *et al*., 2005). PTMs regulate a variety of processes such as gene transcription (Waby *et al*., 2008) and protein degradation (Orford *et al*., 1997); thereby their understanding is critical in the study of cell biology.

1.2.1 Ubiquitination

Ubiquitination is another type of PTM with a role in numerous biological processes (Grabbe *et al*., 2011). One of the best studied is its function in targeting proteins for destruction by the 26S proteasome, a large multisubunit protease. It is known as the ubiquitin-proteasome system (UPS) and is the main proteolytic pathway of eukaryotic cells. It helps the cell to maintain protein homeostasis, a balance between the newly synthesised proteins and degradation of damaged or no longer required proteins, accumulation of which may otherwise be harmful (Hoeller and Dikic, 2009). Not surprisingly, misregulation of protein degradation is associated with many diseases that range from cancer to autoimmune and neurodegenerative disorders (Hershko and Ciechanover, 1998; Martens and Stunnenberg, 2010). Therefore, the UPS is an attractive target for therapeutic intervention and drugs such as bortezomib that inhibits the proteasomal machinery are already used to treat malignancies such as multiple myeloma and mantle cell lymphoma (Shah and Orlowski, 2009; Mohty *et al.*, 2012).

In UPS, the proteins are targeted for degradation by the addition of highly conserved ubiquitin molecules (76 amino acids, ~8.5 kDa). Ubiquitination is a multistep process requiring the activity of three classes of enzymes: an ATP-dependent ubiquitin-activating (E1), ubiquitin-conjugating (E2) and ubiquitin ligase (E3) proteins (**Figure 1.1**; Hershko and Ciechanover, 1998). In the initial step the E1 adenylates the C-terminus of ubiquitin, followed by the attachment of ubiquitin to the E1 via a thioester bond between the carboxyl group of ubiquitin and the thiol group of the active site cysteine (Lee and Schindelin, 2008). The next step involves the transfer of activated ubiquitin to an E2, resulting in the formation of the thioester-linked conjugate. In the final step the E2-ubiquitin complex recruits an E3 to facilitate the transfer of ubiquitin to the target protein. The process culminates in the formation of a stable isopeptide bond between the C-terminal glycine of ubiquitin and the ε-amino group of a lysine residue in the target protein. In humans, ubiquitination involves two E1s, approximately 40 E2 and over 600 ubiquitin ligases (Jackson and Durocher, 2013). Ubiquitination is a reversible modification, where ubiquitin can be removed by about 100 deubiquitinating enzymes (DUBs).

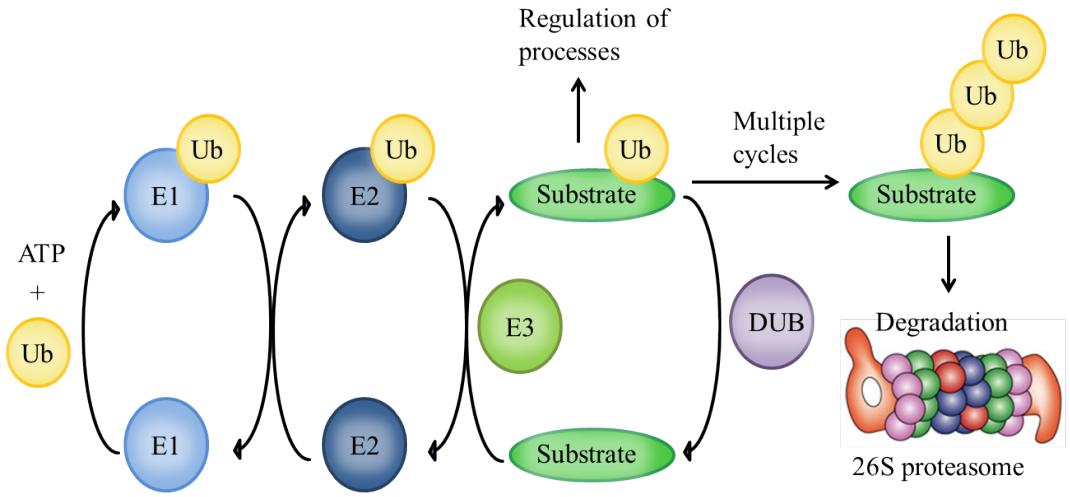

Figure 1.1 The ubiquitin system.
Ubiquitin (Ub) is activated in the ATP-dependent step by ubiquitin-activating enzyme (E1). Activated ubiquitin is transferred to the active site cysteine of an E2 ubiquitin-conjugating enzyme. The E2-ubiquitin complex interacts next with ubiquitin ligase (E3) to facilitate the transfer of ubiquitin to the lysine of the target protein. Multiple cycles can produce polyubiquitin chains that usually label the substrate for degradation by the proteasome. The process can be reversed by deubiquitinating enzymes (DUBs). Figure modified from Jesenberger and Jentsch, 2002.

There are different types of ubiquitination (**Figure 1.2**). For example, the attachment of a single ubiquitin to the substrate is referred to as a monoubiquitination. In other instances, ubiquitin is attached to numerous lysines of the target protein, resulting in multiubiquitination. Furthermore, ubiquitin contains lysine residues within its sequence that can be utilised to construct polymeric chains by the sequential addition of individual ubiquitin moieties (Sadowski and Sarcevic, 2010). Seven internal lysine residues (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, and Lys63) can be used to form ubiquitin chains and their length varies from two to more than ten moieties. In addition, there are branched ubiquitin chains as well as chains containing a mixture of ubiquitin and other ubiquitin-like proteins (Komander and Rape, 2012; Praefcke *et al*., 2012). Different modes of substrate modification adopt distinct structures and physical properties, thereby leading to different cellular outcomes. For instance, monoubiquitination has been shown to regulate DNA repair and gene expression (Passmore and Barford, 2004), while polyubiquitination using Lys48 promotes target-

protein degradation by the proteasome and Lys63 linked chains generally regulate protein-protein interactions (Jackson and Durocher, 2013).

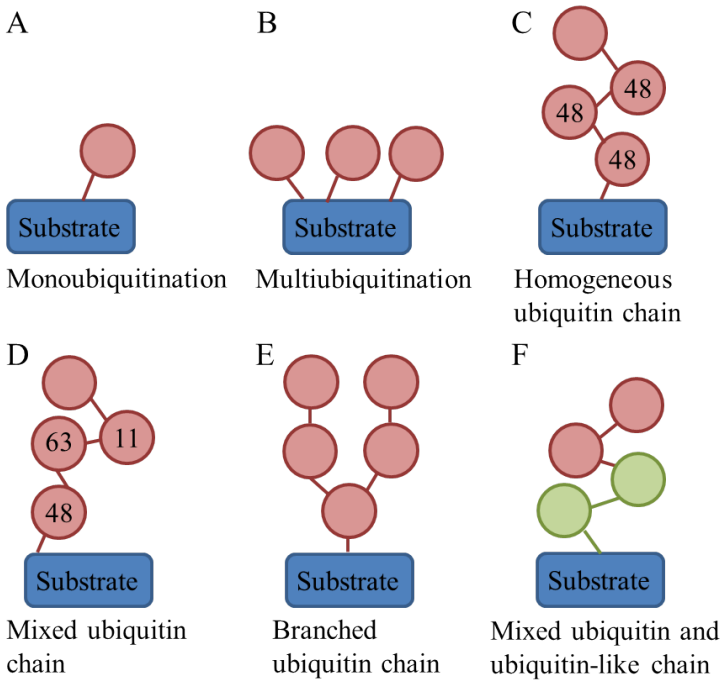

Figure 1.2 The different modes of ubiquitination.
Ubiquitin and ubiquitin-like molecules are shown as red and green circles, respectively. Ubiquitin can be attached to a single (A) or multiple lysine residues (B) of the substrate. The sequential addition using the internal lysine residues of ubiquitin results in polyubiquitination. This produces homogeneous (C) and mixed ubiquitin chains (D). Branched chain (E) is produced when a single ubiquitin is modified with multiple molecules. Some chains contain both ubiquitin and ubiquitin-like proteins (F). Figure adapted from Komander and Rape, 2012.

1.2.2 E3 ubiquitin ligases

The E3 ubiquitin ligases dictate the specificity of ubiquitination by pairing with the substrates that are to be targeted for modification. Not surprisingly, E3s are implicated in a number of malignancies, making them targets for therapeutic intervention (Kirkin and Dikic, 2011). For example, amplification or overexpression of the gene encoding the mouse double minute 2, which regulates the activity of the tumour suppressor p53, facilitates the uncontrolled cell growth in cancer (Marine and Lozano, 2010). Breast cancer gene 1 encodes a tumour suppressor protein with a role in transcription and homologous recombination DNA repair pathway and is often mutated in familial breast

and ovarian cancer (Welcsh and King, 2001). Mutations in the individual Fanconi anaemia proteins with the loss of the multisubunit E3 function lead to Fanconi anaemia and an increased risk of cancer (Moldovan and D'Andrea, 2009). Additional examples of the E3s that are implicated in cancer development and are of a therapeutic interest are discussed in detail in Lipkowitz and Weissman, 2011.

There are two major types of E3s, classified according to their structure and mechanism of action. These are either a homologous with E6-associated protein C-terminus (HECT) domain or a really interesting new gene (RING) domain containing E3 ligases. In addition, a few E3s belong to a distinct U-box family, but are structurally and functionally similar to the RING ligases (Budhidarmo *et al*., 2012).

The HECT family of ligases, which consists of about 30 members in mammals, participates directly in the catalytic process. HECT E3s contain an active-site cysteine residue that accepts ubiquitin from the E2-ubiquitin complex to form an E3-ubiquitin thioester conjugate (Scheffner *et al*., 1995). Ubiquitin is subsequently transferred from an E3-ubiquitin intermediate to the substrate (**Figure 1.3A**). HECT ligases regulate protein trafficking, the immune response and signalling pathways involved in cell growth and proliferation (Rotin and Kumar, 2009). The HECT domain, consisting of around 350 residues is located at the C-terminal end, while the N-terminal domain interacts with the target proteins. The HECT domain can be further subdivided into an N-terminal N-lobe that interacts with the E2 and a C-terminal C-lobe, containing the conserved cysteine (Huang *et al*., 1999). The lobes are joined by a flexible hinge, which brings them together during ubiquitin transfer.

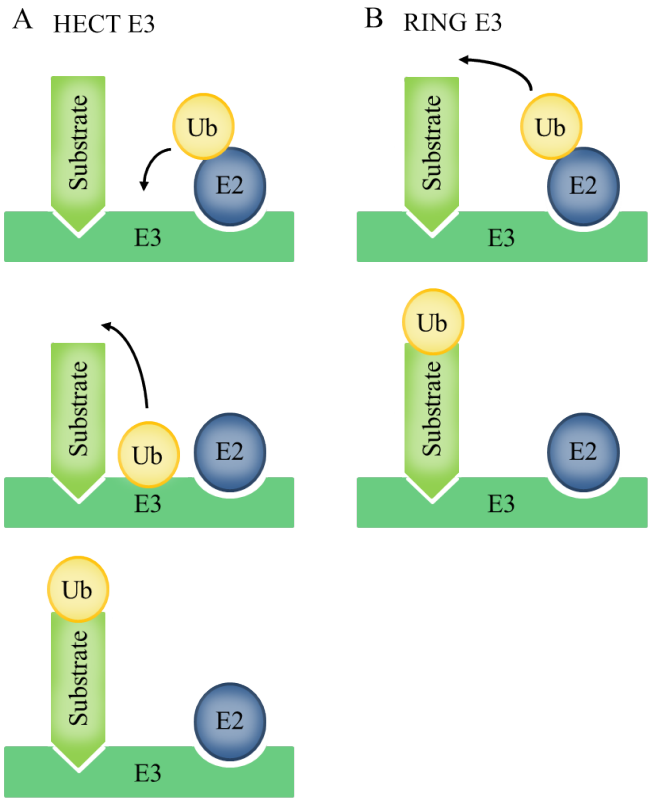

Figure 1.3 Comparison between HECT and RING E3 ubiquitination mechanisms.
(A) The HECT E3s contain a conserved cysteine, which accepts ubiquitin prior to its transfer to the substrate. (B) In contrast, RING E3s act as scaffolds by bringing together E2s and substrates. Figure modified from Deshaies and Joazeiro, 2009.

The majority of E3s belong to the RING domain family, with more than 600 potential RING finger E3s expressed in mammalian cells (Li *et al*., 2008). RING E3 ligases recruit together the E2-ubiquitin intermediate and the target protein to facilitate ubiquitin transfer directly from an E2 to the substrate (**Figure 1.3B**). The precise mechanism remains unclear, but it is believed the catalysis occurs due to the proximity of E2-ubiquitin and substrate (Deshaies and Joazeiro, 2009). Some studies suggest that the RING domain induces a conformational change in E2-ubiquitin to accelerate ubiquitin discharge from the active site of E2 (Seol *et al*., 1999; Skowyra *et al*., 1999). A canonical RING finger comprises 40-60 residues and contains a series of spatially conserved cysteine and histidine residues that coordinate two Zn^{2+} within the interior of the domain (**Figure 1.4A**; Freemont *et al*., 1991). However, unlike the classical DNA-binding C_2H_2 tandem zinc-fingers, two Zn^{2+} in the RING finger are bound in a cross brace arrangement, yielding a compact and rigid α/β fold. The domain is stabilised further by semiconserved residues that form the hydrophobic core. The core region is

made of two β-strands (β1 and β2), one α-helix (α1) and two loops that surround the first and second zinc binding sites, loop 1 and loop 2 (**Figure 1.4B**; Budhidarmo *et al*., 2012). A number of RING finger variants are known. In some the positions of cysteine and histidine residues are swapped, while in others the cysteine is replaced by another zinc coordinating residue such as aspartate (Zheng *et al*., 2002). The U-box E3s adopt a similar tertiary structure to RING E3s, even though the two have low sequence conservation. The U-box E3s contain polar and charged residues that form a network of hydrogen-bonding, thereby replacing the need for the zinc binding sites (Aravind and Koonin, 2000). Members of the RING E3 family can function as monomers, homo- and heterodimers as well as multi-subunit complexes.

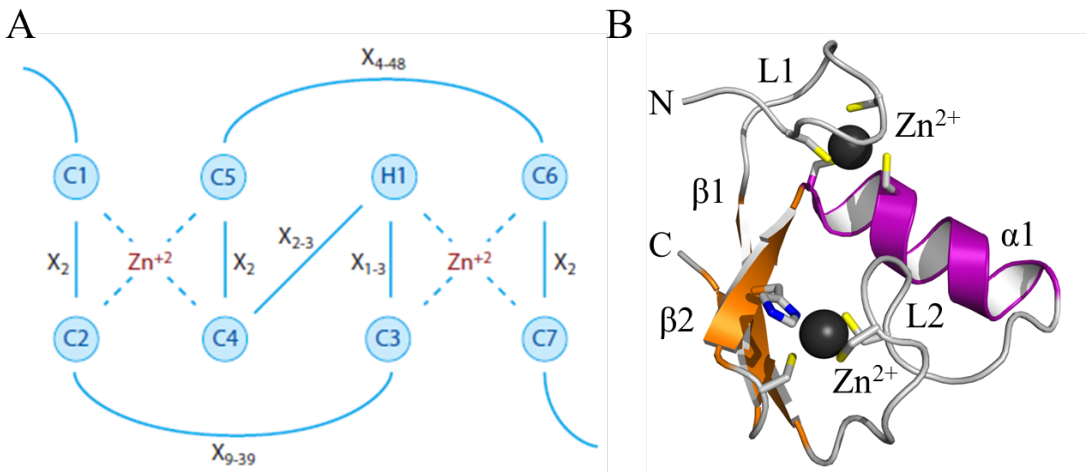

Figure 1.4 The RING finger domain.
(A) The C_3HC_4 RING finger coordinates two Zn^{2+} in a cross brace structure. Cysteine and histidine ligands are denoted with letters C and H, respectively. X_n represents the number of any amino acids in the spacer regions between the zinc ligands. Figure edited from Deshaies and Joazeiro, 2009). (B) Ribbon diagram of the core of the RING domain from Casitas B-lineage Lymphoma protein (PDB code 2y1m; Dou *et al*., 2012). The α-helix and β-strands are coloured in purple and orange, while Zn^{2+} are shown as grey spheres. The zinc ligands are shown as sticks; with the atomic positions coloured: C (grey), N (blue) and S (orange).

1.2.3 The Siah family

Seven-in-absentia homolog (Siah) proteins are evolutionary conserved RING E3 ligases. The first member, seven-in-absentia (SINA) was identified two decades ago as a protein required for the correct development of R7 photoreceptor cells in the

Drosophila eye (Carthew *et al*., 1990 and 1994). Further genetic and biochemical experiments revealed that SINA interacted with Phyllopod and the F-box protein Ebi to mediate the degradation of the transcriptional repressor Tramtrack88 (Ttk), thereby linking SINA function to protein turnover (Tang *et al*., 1997; Li *et al*., 1997; Boulton *et al*., 2000). More recently, a novel *Drosophila* protein was identified that shares 46 % sequence homology to SINA and is able to direct the degradation of Ttk (Cooper *et al*., 2008). SINA homologues (Siah) were first isolated in mice and revealed three murine proteins, Siah1a, Siah1b and Siah2 (Della *et al*., 1993; Holloway *et al*., 1997). In contrast, humans contain two proteins, Siah1 and Siah2. SINA/Siah family proteins are highly conserved across species, except for the N-terminal 40-80 residues. For example, human Siah1 and Siah2 are 97 % and 98 % identical to murine Siah1a/1b and Siah2, and have 69 % and 67 % sequence identity with SINA, respectively. Human Siah proteins are 86 % homologous and 69 % identical with each other. They differ mainly at the N-termini with Siah2 containing an extra 42 residues.

1.2.4 Siah1 substrates

The Siah proteins are localised in both the nucleus and the cytoplasm and control the degradation of a multitude of proteins as diverse as transcriptional regulators, enzymes and neuronal proteins. In total, there are more than 40 known substrates. The high homology between Siah1 and Siah2 is reflected in a large overlap of some target proteins. For example, both proteins play major roles in hypoxia signalling, by regulating the degradation of prolyl hydroxylases (PHDs), factor inhibiting HIF1α (FIH) and homeodomain interacting protein kinase 2 (HIPK2) (Fukuba *et al*., 2008; Nakayama *et al*., 2004; Winter *et al*., 2008). In addition, Siah proteins have member specific substrates. For example, Siah2 is the predominant member that controls the stability of sprouty 2, involved in receptor tyrosine kinase signalling (Nadeau *et al*.,

2007), while only Siah1 can polyubiquitinate the RNA polymerase II elongation factor (ELL2), leading to its proteasomal degradation (Liu *et al*., 2012). Proteins targeted for degradation by Siah1 are listed in **Table 1.1**. Siah proteins mediate ubiquitination either as single proteins or as part of a larger multisubunit complex. For example, Siah1 alone is able to polyubiquitinate β-catenin, a key component of the Wnt signalling pathway controlling cell fate and proliferation, resulting in its degradation (Dimitrova *et al*., 2010). However, this process is more efficient when Siah1 forms an E3 ligase complex with Siah-interacting protein (SIP), Ebi, the adaptor protein Skp1 and adenomatous polyposis coli (APC) (Matsuzawa and Reed, 2001). Therefore, not all Siah1 binding proteins are targeted for destruction. Some binding partners act as cofactor proteins, helping to recruit Siah1's substrates, while others function as negative regulators. Examples of this are the scaffolding proteins Bassoon and Piccolo expressed during neuronal differentiation. They inhibit Siah1's ubiquitinating activity by binding to the RING domain of Siah1 and preventing the interaction with E2 enzymes (Waites *et al*., 2013). Siah1 binding proteins that regulate its activity and are not targeted for degradation are presented in **Table 1.2**. Structurally, little is known about how Siah1 interacts with its binding partners and assembles into a multiprotein E3 ligase complex.

Substrate	Protein function	References
ACK1	A non-receptor tyrosine kinase linked to cell migration, growth and proliferation.	(Buchwald *et al*., 2012)
AF4	Transcription factor with a role in the central nervous system.	(Bursen *et al*., 2004)
BOB1/OBF1	Transcriptional co-activator that regulates B-cell development.	(Boehm *et al*., 2001)
CBP/p300	A lysine acetyl-transferase that control cell growth, transformation and development.	(Grishina *et al*., 2012)
c-myb	Transcriptional activator controlling cellular proliferation and differentiation.	(Tanikawa *et al*., 2001)
CtIP	Endonuclease involved in DNA repair and cell cycle control.	(Germani *et al*., 2003)
ELL2	A component of the super elongation complex required to increase the catalytic rate of RNA polymerase II.	(Liu *et al*., 2012)
N-CoR	Transcriptional co-repressor that promotes chromatin condensation.	(Zhang *et al*., 1998)
PML	Transcriptional co-regulator found in PML-nuclear bodies and mediates tumour suppression, apoptosis and DNA damage response.	(Fanelli *et al*., 2004)
TIEG1	Transcription factor involved in the regulation of cell growth.	(Johnsen *et al*., 2002)
FIH	A hydroxylase that functions as an oxygen sensor.	(Fukuba *et al*., 2008)
PHD proteins	Enzymes involved in various hypoxia-influenced processes.	(Nakayama *et al*., 2004)
Synphilin-1	A cytoplasmic protein with the role in synaptic function and protein degradation and in the pathogenesis of Parkinson's disease.	(Liani *et al*., 2004)
Synaptophysin	A protein of the presynaptic vesicle exocytosis machinery.	(Wheeler *et al*., 2002)
Group1 mGluRs	G protein-coupled receptors involved in many aspects of normal brain function.	(Ishikawa *et al*., 1999)
β-catenin	Transcription factor mutated in multiple cancers.	(Liu *et al*., 2001; Dimitrova *et al*., 2010)
BAG-1	Protein involved in pathways controlling cell proliferation and migration.	(Matsuzawa *et al*., 1998)
DCC	Tumour suppressor protein with a role in mediating directional migration in the developing nervous system.	(Hu and Fearon, 1999)
EB3	Microtubule plus-end binding protein that controls microtubule and spindle dynamics.	(Ban *et al*., 2009)
HIPK2	Protein kinase involved in regulation of transcription and apoptosis.	(Winter *et al*., 2008;
Kid	A nuclear protein involved in spindle formation and the movements of chromosomes during cell division.	(Germani *et al*., 2000)
KSHV ORF45	Human herpesvirus 8 encoded ORF45 protein is essential for viral infection.	(Abada *et al*., 2008)
Numb	A signalling adapter protein involved in neurogenesis.	(Susini *et al*., 2001)

OGHDC	A rate-limiting enzyme in the mitochondrial Krebs cycle.	(Habelhah *et al*., 2004)
PEG10	A mediator of apoptosis.	(Okabe *et al*., 2003)
PLC_{ε}	Enzyme that converts polyphosphoinositides into the second messengers.	(Yun *et al*., 2008)
Polycystin-1	The integral membrane protein involved in renal tubulogenesis.	(Kim *et al*., 2004)
Tramtrack88	Transcriptional repressor involved in eye development of *Drosophila*.	(Li *et al*., 1997; Tang *et al*., 1997)
TRB3	A putative kinase that down regulates various signal transducers.	(Zhou *et al*., 2008)
HPH2	A subunit of a large multimeric polycomb complex 1 implicated in the maintenance of transcriptional repression of target genes.	(Wu *et al*., 2010)
T-STAR	Alternative splicing factor.	(Venables *et al*., 2004)

Table 1.1 Siah1 substrates that are targeted for proteasomal degradation.
Abbreviations: ACK1, activated Cdc42-associated kinase 1; BOB1, B-cell Oct-binding protein; CBP, CREB binding protein; CtIP, C-terminal interacting protein; DCC, deleted in colorectal cancer; EB3, end-binding protein 3; ELL2, RNA polymerase II elongation factor; FIH, factor inhibiting HIF1α; HIPK2, homeodomain interacting protein kinase 2; HPH2, the *Homo sapiens* polyhomeotic homologue 2; Kid, Kinesin like DNA binding protein; KSHV ORF45, Kaposi's sarcoma-associated herpes virus open reading frame 45 protein; mGluRs, metabotropic glutamate receptors; N-CoR, nuclear receptor co-repressor; OBF1, Oct binding factor 1; OGHDC, 2-oxoglutarate dehydrogenase complex; PEG10, paternally expressed gene 10; PHD, prolyl-hydroxylases; PLC_{ε}, phospholipase C epsilon; PML, promyelocytic leukaemia protein; TIEG1, transforming growth factor β inducible early gene-1; TRB3, tribbles homolog 3.

Interacting protein	Protein function	References
UbcH5, UbcH8 and UbcH9	E2 enzymes that accept ubiquitin from the E1 complex and attach it covalently to other proteins.	(Matsuzawa *et al*., 2001; Wheeler *et al*., 2002; Hu *et al*., 1997)
α-synuclein	Its function is unknown.	(Liani *et al*., 2004)
Dab-1	Signal transducer functioning in neural development.	(Park *et al*., 2003)
α-tubulin	Cytoskeletal protein.	(Germani *et al*., 2000)
APC	Tumour suppressor that promotes β-catenin degradation together with Siah1.	(Liu *et al*., 2001)
GAPDH	Glycolytic protein that participates in transcription, RNA transport and apoptosis.	(Hara and Snyder, 2006)
PEG3	A possible transcription factor that induces apoptosis in cooperation with Siah1.	(Relaix *et al*., 2000)
Phyllopod	*Drosophila* adapter protein involved in cell differentiation and organ development.	(Li *et al*., 2002)
POSH	A scaffold protein of the apoptotic JNK pathway.	(Xu *et al*., 2006)
SIP	Adapter protein involved in β-catenin degradation together with Siah1.	(Matsuzawa *et al*., 2001)
TRF2	A component of telomere DNA-interacting complex with a role in telomere maintenance.	(Fujita *et al*., 2010)
Bassoon	A scaffolding protein involved in organization of synaptic active zones.	(Waites *et al*., 2013)
Piccolo	A scaffolding protein that maintains synapse integrity together with Bassoon by regulating protein ubiquitination and degradation.	(Waites *et al*., 2013)
EEF1D	Elongation factor inhibiting Siah1 ubiquitin ligase activity.	(Wu *et al*., 2011)

Table 1.2 Siah1-interacting proteins that are not targeted for proteasomal degradation.
Abbreviations: APC, adenomatous polyposis coli; Dab-1, disabled-1; EEF1D, Eukaryotic translation elongation factor 1 delta; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; PEG3, paternally-expressed gene 3; POSH, plenty of SH3s; SIP, Siah interacting protein; TRF2, telomeric repeat-binding factor 2; Ubc, ubiquitin-conjugating enzyme.

1.2.5 Siah1 and disease

Siah proteins play major roles in a number of important signalling pathways, including Ras, hypoxia and DNA damage (Schmidt *et al*., 2007; Nakayama *et al*., 2004; Liu *et al*., 2001). These pathways are often misregulated in cancer, but the exact function of Siah family members in tumour progression remains disputed. They were reported to be both oncogenic and tumour suppressive, with the initial experiments suggesting Siah1 may

act as a tumour suppressor. *In situ* hybridisation of cancerous tissue using microarrays revealed reduced expression levels of both Siah genes in breast cancer, and these correlated significantly with prognostic factors and clinical outcome (Confalonieri *et al*., 2009). Immunofluorescence microscopy studies showed that Siah1 protein levels were decreased in human tumour tissue compared to normal controls, suggesting Siah1 is either a tumour suppressor or its loss is the outcome of tumour progression (Bruzzoni-Giovanelli *et al*., 2010). Studies in hepatoma cells revealed that low levels of Siah1 correlated positively with reduced apoptosis, while over-expression of Siah1 in 293 epithelial cells and GM701, an SV40-transformed human fibroblast cell line led to growth arrest (Yoshibayashi *et al*., 2007; Matsuzawa *et al*., 1998). Further studies observed enhanced radiosensitivity of human breast cancer cells upon overexpression of Siah1 (He *et al*., 2010). The above mentioned studies allude to a tumour suppressive function of Siah1. However, proteins acting as tumour suppressors are often mutated in cancer, but currently there is little evidence of that for Siah1. Two missense mutations, occurring at a low frequency in gastric cancer have been reported so far, resulting in stabilisation of oncogene β-catenin and apoptosis block, while no mutations were identified in another study looking at other cancer types (Kim *et al*., 2004; Medhioub *et al*., 2000). A more recent study observed that nuclear accumulation of Siah1 enhanced proliferation and migration of hepatocellular carcinoma cells, suggesting Siah1 is pro-tumourigenic (Brauckhoff *et al*., 2011). In addition, a number of studies that did not differentiate between Siah family members were performed using the animal cancer models and their data uniformly support an oncogenic role of Siah proteins (Schmidt *et al*., 2007; Möller *et al.*, 2009). Together these studies depict an opposing mix of results and additional experiments are required to establish the role of Siah1 in tumour progression.

1.2.6 Siah1 as a drug target

Given the importance of Siah proteins as master regulators of different signalling pathways and their relevance in cancer development, they may be useful therapeutic targets. Siah proteins contain an N-terminal RING domain, two zinc finger motifs and a C-terminal substrate binding domain (SBD) (**Figure 1.5A**). They are dimeric proteins, with the region in SBD involved in dimerisation. Three structures of Siah1 SBD without the RING domain have been determined, with the highest resolution of 2.20 Å (**Figure 1.5B**). Siah proteins have three main sites for inhibition, resulting in potentially different biological effects. These are the disruption of protein-protein interactions between Siah and either substrate, E2 enzyme or another Siah subunit in a dimer. The first inhibitor identified to target Siah function was the PHYL peptide. This peptide, containing a 23 residue sequence from phyllopod protein is a potent binder of Siah SBD and was shown to reduce the tumour growth and metastasis in a number of models (House *et al*., 2003; Möller *et al*., 2009; Qi *et al*., 2010). Inhibitors targeting Siah SBD may be specific for these proteins, but they may not distinguish between the two Siah isoforms due to their high homology in this region. The N-terminal RING domain of Siah is essential for their ubiquitinating activity. RING domain mutants of Siah abrogate substrate degradation and impair with Ras signalling pathway, thereby interfering with cancer progression (Hu and Fearon, 1999; Wong and Möller, 2013). Inhibitors may be designed to target N-terminus/RING domain to block the binding of E2. The N-terminal regions vary between Siah1 and Siah2; thereby isoform selective inhibitors may be designed (Della *et al*., 1993). However, RING domain inhibitors may also impair with the function of other RING-containing E3 ligases. A third method to interfere with Siah activity would be to disrupt its dimerisation interface. Protein Zyxin involved in the regulation of actin cytoskeleton dynamics and cell motility was identified recently to inhibit Siah1 homodimerisation (Crone *et al*., 2011). However, further studies are

needed to elucidate the impact of Siah homo- and heterodimerisation and the outcome of disrupting such interactions in cancer. Proof of concept for small molecule inhibitors targeting Siah was obtained with menadione (vitamin K12). It was initially identified to inhibit Siah2 in a screen of U.S. Food and Drug Administration approved therapeutic drugs using electro-chemiluminescent-based assay (Shah *et al*., 2009). Further *in vitro* based dose-response experiments looking at Siah self-ubiquitination also showed the inhibition of Siah1. The same paper reports that menadione blocked the Ras/mitogen-activated protein kinase pathway, attenuated hypoxia and inhibited melanoma tumourigenesis. Although the inhibitory mechanism of menadione is not known, this study supports the rationale for screening for inhibitors of Siah.

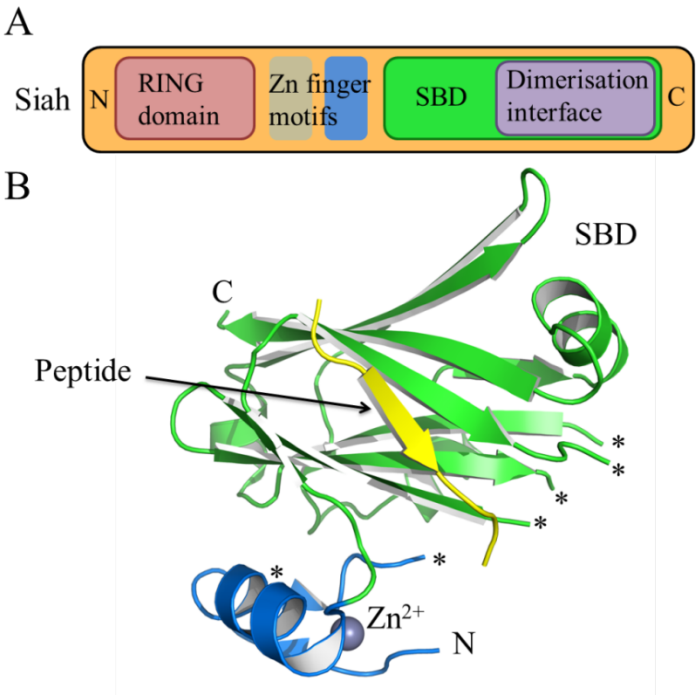

Figure 1.5 Structure of Siah.
(A) Schematic diagram shows the domain architecture of Siah. The main difference between Siah1 and Siah2 is in the length of the N-terminal sequence prior to the RING domain. (B) The structure of Siah1 SBD determined to 2.20 Å (PDB code 2a25; Santelli *et al*., 2005). SBD and the adjacent zinc finger motif are coloured in green and marine, respectively. Zn^{2+} is shown as a grey sphere and the disordered loops are marked with asterisks. A 13 residue peptide containing a conserved VxP motif present in many substrates of Siah is shown in yellow.

1.3 SUMOylation

In addition to ubiquitin, there are a number of ubiquitin-like proteins that can be conjugated to the target proteins. One of them is small ubiquitin-like modifier (SUMO), which was first observed to modify the nucleoporin Ran GTPase-activating protein 1 (RanGAP1) leading to its translocation (Matunis *et al*., 1996; Mahajan *et al*., 1997).

Later studies revealed that SUMO participated in a wide variety of cellular processes including cell cycle control, gene transcription, cellular localisation, degradation and chromatin organisation (Müller *et al*., 2001; Seeler and Dejean, 2003; Verger *et al*., 2003). However, its primary molecular role is to regulate interactions of the modified substrates with other proteins. SUMO (~100 residues, ~8 kDa) shares only 18 % sequence identity with ubiquitin, but structural studies by nuclear magnetic resonance (NMR) revealed the overall structure closely resembles that of ubiquitin, characterised by a tightly packed globular fold with β-sheets wrapped around one α-helix (Bayer *et al*., 1998; Müller *et al*., 2001). Lower eukaryotes such as yeast contain a single isoform of SUMO, while mammals have three homologues (SUMO-1, -2 and -3), in contrast to a single type of ubiquitin. SUMO-1 is 50 % identical to SUMO2/3, while the later share 95 % homology and appears to be functionally redundant (Johnson, 2004). Analysis of the protein sites modified by SUMO, revealed that conjugation occurs on the lysine residue present in the ψKxE sequence, where ψ corresponds to a large hydrophobic amino acid and x represents any amino acid (Rodriguez *et al*., 2001). This consensus motif is also found at the N-terminus of SUMO2/3, allowing them to form polySUMO chains. The motif is absent from SUMO-1, thereby it is unable to form polymers and functions as a polySUMO chain terminator (Kroetz, 2005; Ulrich, 2009). The SUMO conjugation mechanism is similar to the ubiquitination pathway. It requires ATP and a cascade of three enzymes. SUMO-activating enzyme (E1) activates the SUMO C-terminal carboxyl group followed by formation of a thioester conjugate. The SUMO is subsequently transferred to the active site cysteine residue of the SUMO-conjugating enzyme (E2). The final transfer of SUMO to the target protein is generally mediated by SUMO ligases, which bring E2-SUMO complex and substrate together, thus enhancing SUMOylation (Ulrich, 2009). In contrast to a large variety of enzymes participating in

ubiquitination process, SUMOylation uses only a single E1, one E2 and about ten E3s (Yeh, 2009). SUMOylation can be reversed by SUMO-specific proteases (SENPs).

1.3.1 The SENP family

SENPs perform two functions; deconjugate SUMO moieties from the modified proteins and catalyse the maturation of SUMO precursors. The SUMO genes produce precursor proteins, which are subsequently processed by SENPs into a mature form by cleaving the peptide bond to expose the C-terminal diglycine motif (**Figure 1.6A**). This form of SUMO is then linked via its carboxyl group with the ε-amino group of a lysine in the substrate, forming an isopeptide bond. This bond can be later cleaved by the isopeptidase activity of SENPs to release the target protein and SUMO back into the SUMO conjugation cycle (**Figure 1.6B**). Thus, SENPs initiate the SUMOylation cycle by regulating the availability of free SUMO and also control the modification status of individual substrates. SENPs belong to the C48 protease family that shares a conserved catalytic domain with a characteristic catalytic triad of histidine, aspartate and cysteine (Yeh *et al*., 2000). Humans contain six SENPs, which can be divided into three groups. The first group contains SENP1 and SENP2, which are involved in both precursor processing and deconjugation of all three SUMO isoforms. SENP3 and SENP5 belong to the second group. They are found in the nucleolus and appear to specialise in the proteolysis of SUMO2/3 (Hickey *et al*., 2012). The third group includes SENP6 and SENP7, which are localised in the nucleoplasm and involved in the editing of polySUMO2/3 chains (Hattersley *et al*., 2011; Shen *et al*., 2009).

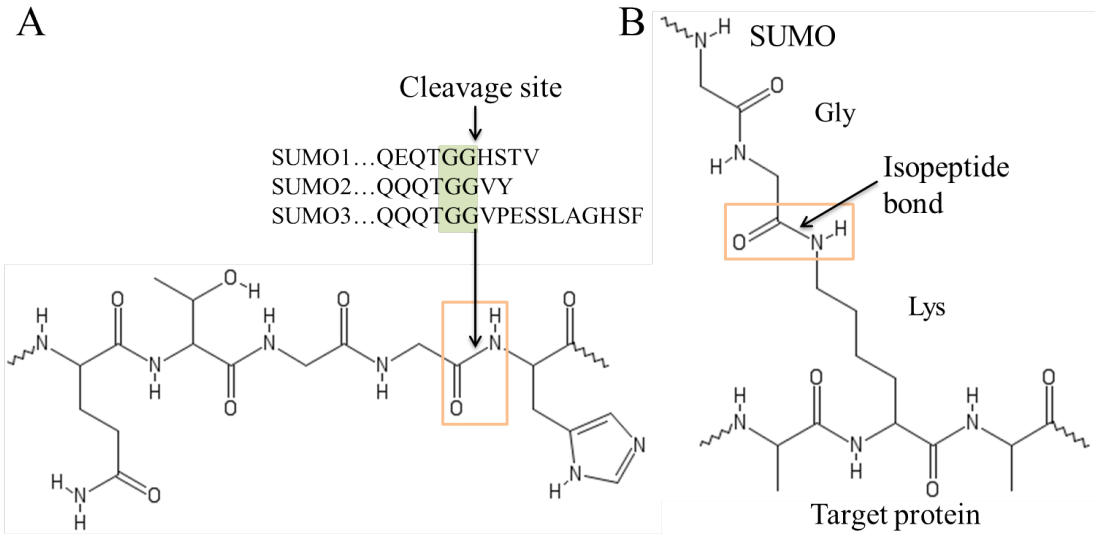

Figure 1.6 Isopeptidase and endopeptidase activities of SENPs.
(A) The cleavage at the C-terminal end of SUMO results in its mature form. Diagram shows the C-termini of three mammalian SUMO isoforms, with the diglycine motif coloured in green. The chemical structure of the C-terminal end of SUMO1 is presented, with the site of cleavage marked within a box and the scissile bond being cleaved indicated by an arrow. (B) SUMO deconjugation from the target protein. The chemical structure of the isopeptide bond between target lysine and SUMO is shown. The site of cleavage is indicated with a box and an arrow points to the isopeptide bond.

1.3.2 SENP1

SENP1 has a broad specificity, processing three pro-SUMO isoforms to their mature form and deconjugating them from modified proteins (Xu and Au, 2005; Gong *et al*., 2000). It contains the C-terminal catalytic domain and the N-terminal localisation region (**Figure 1.7A**). The later contains a nuclear localisation signal, targeting SENP1 to the nucleoplasm (Gong *et al*., 2000). Crystal structures of the apo-SENP1 and SENP1 in complex with either pro-SUMO1 or SUMO1-modified RanGAP1 reveal its catalytic region and the mechanism of action (Shen *et al*., 2006). SENP1 shares a common fold with other SENPs, in which the core is made of a five stranded mixed β-sheet situated adjacent to two α-helices (**Figure 1.7B**). The conserved His533 and Asp550 are situated on two neighbouring β-stands, with Cys603 located on a central α-helix. Analysis of the structure of SENP1 bound to SUMO1-RanGAP1 indicates the minimal recognition of the substrate, with the main interface area formed between the conserved C-terminal

strands of SUMO1 and the main chain of SENP1. Structural studies also suggest that the binding of SUMO precursors and SUMO-target conjugates to SENP1 results in the opening of a tunnel to accommodate diglycine motif (Yeh, 2009). This tunnel is formed by two tryptophan residues and is too narrow to fit residues other than glycine. Closing of the tunnel then leads to *trans, cis*-isomerisation of the amide nitrogen atoms of the scissile bond. The resulting *cis* peptide bond is thermodynamically unfavoured and promotes cleavage (Hickey *et al*., 2012).

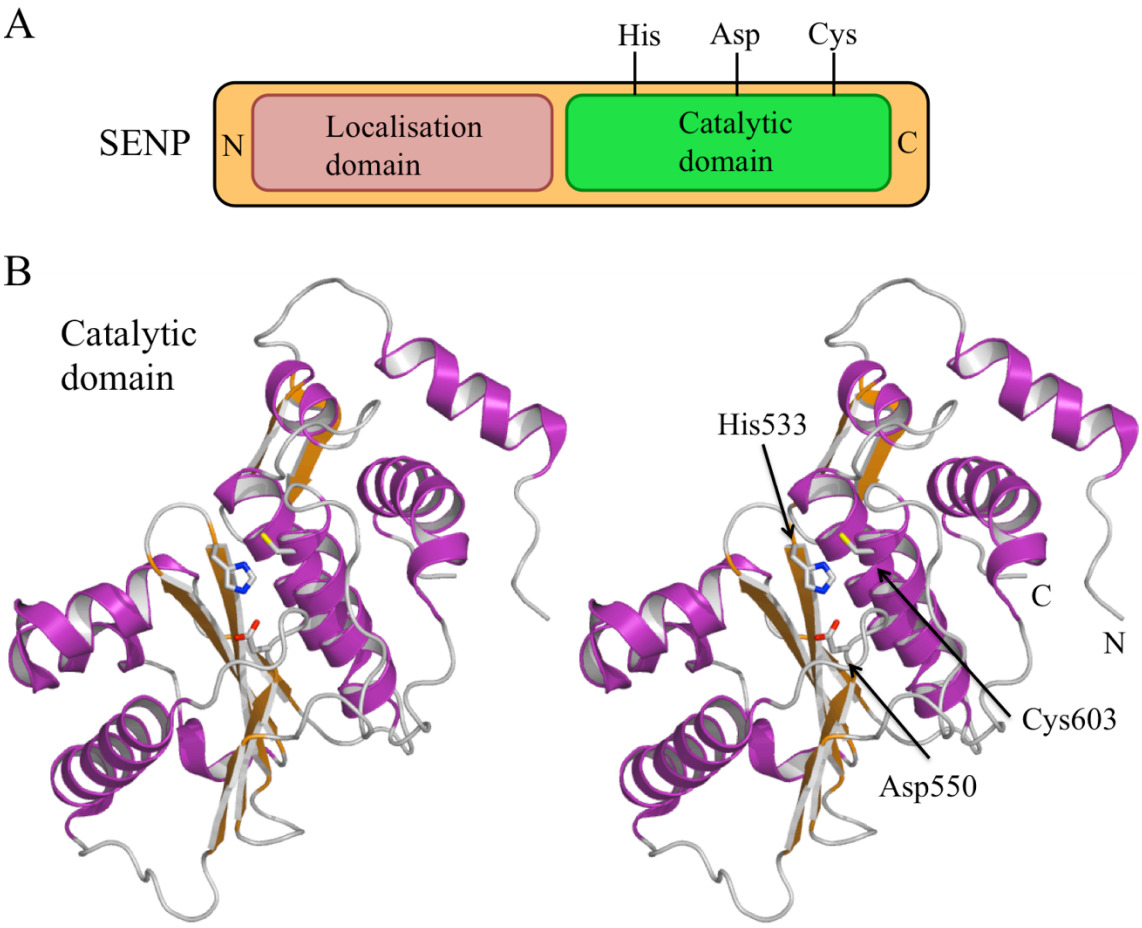

Figure 1.7 Structure of SENP1.
(A) Domain organisation of SENP1. (B) The stereo image is showing the catalytic domain of SENP1 determined to 2.45 Å (PDB code 2iyc; Shen *et al*., 2006). The α-helices and β-strands are coloured in purple and orange, respectively. A catalytic triad is shown as sticks, with the atomic positions coloured: C (grey), N (blue), S (yellow) and O (red).

SENP1 is overexpressed in prostatic intraepithelial neoplasia and prostate cancer lesions, and promotes androgen receptor dependent transcription and cell proliferation

(Bawa-Khalfe *et al*., 2007; Kaikkonen *et al*., 2009). Silencing of SENP1 attenuates the expression of several AR target genes and reduces androgen-stimulated growth of LNCaP cells (Kaikkonen *et al*., 2009). These data suggest that SENP1 could play an important role in prostate carcinogenesis. Development of small synthetic inhibitors is currently underway to validate SENP1 as a potential drug target (Uno *et al*., 2012; Chen *et al*., 2012; Sommer *et al*, 2013; Madu *et al*., 2013).

1.4 Introduction to fragment screening

1.4.1 Fragment based drug discovery

A typical collection of compounds employed for high throughput screening (HTS) in the 1990s consisted of molecules with at least 15 heavy atoms (Tounge and Parker, 2011). The identified hits had to be subsequently optimised to include the desirable absorption, distribution, metabolism, excretion and toxicity properties (Jhoti, 2005). However, as the size of the initial hits was already similar to the common drug molecules, the scope for optimisation was limited and these compounds often failed at the later stage of drug development.

A survey, looking at the physico-chemical properties of known orally available drugs, resulted in a set of parameters known as the Lipinski rule of five, which improved the likelihood of compounds being membrane permeable and easily absorbable by the body (**Table 1.3**; Lipinski *et al*., 2001). These rules were widely implemented due to their simplicity to assess not only the oral bioavailability, but also overall drug-likeness of compounds. Although some drugs did not obey the rules, they still raised the awareness for such properties (Wenlock *et al*., 2003). In particular, the lead-like molecules were re-evaluated with a view of retaining their properties during the later stages of

optimisation. This meant setting tighter restrictions on their chemical properties, such as limiting their molecular weight to <400 Da (Leach and Hann, 2011).

Property	Lipinski rule of five	Rule of three
Mw	<500	<300
H-bond donors	≤5	≤3
H-bond acceptors	≤10	≤3
ClogP	=5	=3

Table 1.3 Comparison of Lipinski rule of five to rule of three.
Data taken from Lipinski *et al*., 2001 and Congreve *et al*., 2003.

More recently, a new approach of using even smaller molecules termed fragments became popular. Fragments have lower molecular weight and generally contain fewer functional groups, with most fragment libraries complying with properties outlined by the rule of three (**Table 1.3**; Congreve *et al*., 2003). Fragments contain a number of unique properties that distinguish them from typical HTS compounds. Firstly, fragments have higher probability of binding due to fewer interaction constraints. For example, many proteins contain cavities, which are able to accommodate a simple structure like a phenol ring with single hydrogen bond donor. However, a more specific binding site is needed to fit a compound with three hydrogen bond contacts while maintaining good van der Waals contacts and ligand geometry (Tounge and Parker, 2011). Therefore, the likelihood of compound binding goes up as its complexity goes down (Hann *et al*., 2001). Evidence of this can be observed in the high ligand efficiencies detected for fragments, which are at least as high as those for larger hit-molecules (Bembenek *et al*., 2009; Nissink, 2009). Ligand efficiency (LE) measures the binding affinity per heavy atom (non-hydrogen) (Hopkins *et al*., 2004):

$$LE = \frac{\Delta G_{binding}}{\#HA} = \frac{-RTlnK_{\mathrm{d}}}{\#HA}$$

where *#HA* is the number of heavy atoms, $\Delta G_{binding}$ (Gibbs free energy), *R* the gas constant, *T* the absolute temperature and K_d the dissociation constant.

Secondly, drug-like compounds of HTS size (350-550 Da) can theoretically produce over 10^{60} unique combinations, but an estimated number of possible molecules composed of 11 or fewer heavy atoms is about 10^7 (Bohacek *et al*., 1996; Ertl, 2003). Therefore, a smaller library of about 10^3 fragments can explore a significantly larger volume of chemical space compared to the usual HTS collection of 10^6 larger compounds. In addition, smaller libraries reduce the screening and compound synthesis time. Fragments are also likely to show less prejudice for specific target classes, as many corporate HTS collections contain compounds with tendency to hit kinases for example (Tounge and Parker, 2011), although criteria used in library design will always influence their final applicability.

1.4.2 Fragment screening

Fragments generally have binding affinities for the target protein in the range of 0.1-10 mM. Detection of such weak binding requires sensitive screening methods. In addition, the screen is performed at higher concentrations of fragments than those employed in typical HTS biochemical assays. **Figure 1.8** depicts the range of methods available for fragment screening. The screening methods can be divided into two groups. Traditional biophysical methods such as Surface Plasmon Resonance (SPR) and Differential Scanning Fluorimetry (DSF) only detect binding, while the second group that includes protein-detection NMR and X-ray crystallography offer further information about the binding site. Most screening techniques provide various degree of affinity information. Traditional techniques have a high throughput format and are often used in the initial fragment screen, which can be later followed up by the structural method to reveal binding interactions. In all fragment based drug discovery studies, structural

information is desirable to progress hits towards compounds with higher efficacy and selectivity, known as leads. The screening techniques are discussed in more detail in the later chapters.

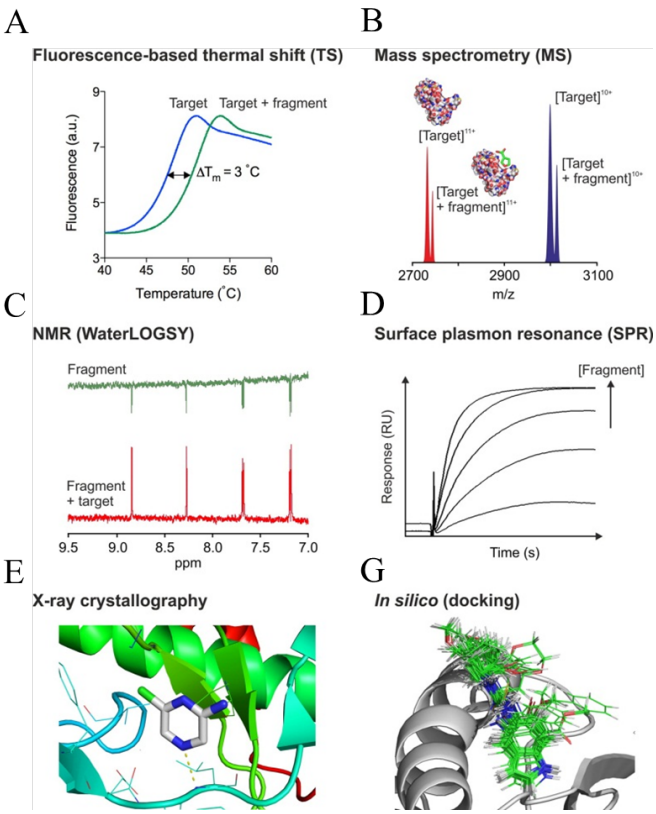

Figure 1.8 Examples of various fragment-screening techniques. (A) Thermal shift assay monitors changes in the melting temperature due to ligand binding. (B) Mass spectrometry is used to detect protein-ligand interactions directly. (C) 1D NMR is generally used for screening and 2D NMR for further hit characterisation. (D) SPR sensorgrams depict a fragment titration experiment. (E) X-ray crystallography can provide information about the fragment binding site. (G) Virtual screening is a computational method of fragment screening. Figure adapted from Scott *et al*., 2012.

Validated fragment hits are optimised to enhance their potency using rational design coupled with the cycles of structural studies and affinity data measurements (Scott *et al*., 2012). Approaches employed for fragment optimisation fall into three groups: fragment merging, linking and growing. Fragment merging involves the incorporation of different regions of overlapping molecules into a single compound. In fragment linking, the separate molecules binding at non-overlapping sites are joined together. Fragment growing is a frequently used method, where molecules are modified using chemical synthesis to include additional binding contacts (Hubbard, 2008).

1.5 Aims

Siah1 and SENP1 represent crucial regulators of a number of signalling pathways and have been shown to facilitate cancer progression and spread (Wong and Möller, 2013; Yeh, 2009). Therefore, they could be attractive targets for therapeutic intervention. Development of inhibitors is currently underway, with a number of virtual screening studies reporting inhibitors of SENP1 (Madu *et al*., 2013). Long inhibitory peptides mimicking the substrate have been reported for Siah1 and demonstrated the desirable effect in cells (House *et al*., 2003). These early compounds provide a proof-of-concept for screening for novel inhibitors with better efficacies.

Fragment screening has emerged as a complementary and contrasting approach to HTS to identify the starting leads. There are a number of drug candidates developed using this method that are currently in the clinical trials, thereby supporting the power of this method (Scott *et al*., 2012). The methodology proved successful against the initial targets such as kinases and phosphatases, and it is now being diversified to include more challenging classes of targets. The main aims of the work described in this thesis were to assess the applicability of the fragment screening for targeting protein-protein interactions and cysteine proteases such as Siah1 and SENP1, respectively. Protein-protein interactions generally differ from protein-ligand binding sites. They form an extensive interface with multiple contacts in contrast to a single profound cavity found in most protein-ligand active sites. Analysis of the contributions of individual residues in protein-protein interactions revealed the free energy of binding is not distributed evenly and observed the hot spots enriched in tryptophan, tyrosine and arginine (Bogan and Thorn, 1998). A number of studies discovered small molecule inhibitors with drug-like potencies that targeted these hot spots (Wells and McClendon, 2007). However, further studies are required to assess protein-protein interactions that involve other

protein folds and contain a different set of interface residues. There are a number of SENP inhibitors available, but more potent and isoform selective inhibitors are required (Madu et al., 2013). SENP1 might be amenable to a fragment-based approach to identify molecules targeting its shallow active site and those that distinguish the active site's unique features.

Another objective was to probe the characteristics of different screening methods. Any fragments identified in the process would be advantageous for elucidating the roles of these proteins in post-translational processes as well as validating them as therapeutic targets, and might even serve as starting points for drug-development in the future. Two fragment libraries were screened using DSF and SPR methods. The outcomes of these fragment screens and the limitations of different screening methods are reported and discussed in chapters three and four for Siah1 and SENP1, respectively.

Siah1 is a scaffolding protein interacting with its target proteins to facilitate their ubiquitination and subsequent degradation (**Table 1.1**). It also interacts with many non-substrate proteins that regulate its activity and sometimes forms the multiprotein E3 ligase complexes to mediate ubiquitination of certain substrates (**Table 1.2**; Matsuzawa and Reed, 2001). A number of substrates have been shown to bind Siah1 via a conserved VxP motif and a few NMR studies analysed the interaction sites using short protein fragments (House *et al*., 2006; Santelli *et al*., 2005). However, this information is absent for majority of Siah1-binding proteins and no structural data are available to understand the formation of Siah1 ligase complexes. Furthermore, some published studies reporting the novel substrates of Siah1 are not conclusive and additional experiments are needed to validate them. Chapter five covers protein interaction studies. A number of previously reported Siah1-binding proteins were evaluated with the aim to elucidate the precise binding mechanism.

Chapter 2

Materials and methods

2.1 General materials

2.1.1 Reagents

General reagents were purchased from suppliers Sigma-Aldrich, BDH Chemicals and Fisher unless otherwise stated. Tobacco etch virus (TEV) protease with a polyhistidine-tag was prepared by Keri Barrack (University of Dundee) and stored at -80 ^{o}C until required. 1 mg of TEV protease was used per 20 mg of protein in a cleavage reaction, which was performed at room temperature for three hours. NuPAGE 10 % Bis-Tris and Mini-PROTEAN TGX precast gels as well as premixed electrophoresis running buffers were purchased from Invitrogen and Bio-Rad, respectively. DNA and protein ladders were HyperLadder I (Bioline) and SeeBlue Plus2 Pre-stained Standard (Invitrogen). Coomassie Brilliant Blue and InstantBlue (Expedeon) were used to stain the protein gels.

2.1.2 Bacterial strains and growth media

TOP 10 competent cells (Invitrogen) and XL1 blue competent cells (Stratagene) were used for DNA preparation. *Escherichia coli* Bl21 (DE3) Gold cells (Novagen) were used for expression of all protein constructs. Luria-Bertani (LB; Bertani, 1951) broth and LB-agar plates were purchased from the College of Life Sciences Media Kitchen (University of Dundee). Super optimal broth with catabolite repression (S.O.C.) medium (Invitrogen) was used for bacterial transformations.

2.1.3 Peptide inhibitors

Two synthetic peptides were purchased from commercial sources to be used as the positive control in the SPR fragment screen (**Table 2.1**).

Sequence	Molecular weight	Purity	Commercial source
Ala-Ser-Leu-Gln-Arg-Val-Arg-Arg-Pro-Val-Ala-Met-Val-Met-Pro-Ala-Arg-Arg-Thr-Pro-His-Val-Gln (plectin-1$^{95-117}$ peptide)	2697 Da	>95 %	Cambridge Research Biochemicals
Arg-Pro-Val-Ala-Ala-Val-Arg-Pro-Thr	966 Da	>98 %	GenScript

Table 2.1 Peptide inhibitors.

2.2 Molecular cloning

2.2.1 Protein constructs and expression plasmids

The proteins used for this work (unless mentioned otherwise) are given in the **Table 2.2**. The gene encoding SENP1 in a pHISTEV30a vector was a kind gift of Ron Hay (University of Dundee). The gene fragments encoding Siah1, SIP and PEG3 with flanking restriction sites were purchased in the pUC57 vector (GenScript). These were subsequently cut out using appropriate restriction enzymes and ligated into the pET15b-TEV vector. Thomas Eadsforth (University of Dundee) cloned the *Bc*CCP gene from genomic DNA into the pET15b-TEV vector. The pET15b-TEV and pHISTEV30a plasmids encode genes conferring resistance to carbenicillin and kanamycin, respectively. Both plasmids are designed to express the protein of interest with an N-terminal hexa histidine tag and a TEV protease cleavage site.

Protein	Residues	UniProt No.
SENP1	415-644	Q9P0U3
Siah1	91-282	Q8IUQ4
SIP	1-228	Q9HB71
PEG3	40-130	Q9GZU2
*Bc*CCP	1-384	B4EEQ5

Table 2.2 Protein constructs.

2.2.2 DNA manipulation

Other gene fragments encoding alternative, shorter protein constructs were generated using polymerase chain reaction (PCR). The technique involves repeated cycles of heating and cooling to amplify the desired DNA sequence. The heating is required to separate the DNA strands, while the cooling allows the primers to anneal, before the extension of the primed strands can proceed. Synthetic genes, mentioned in section 2.2.1, were used as DNA templates to generate shorter gene fragments. Primers were purchased from Thermo Fisher Scientific. The PCR reactions were carried out using parameters outlined in **Table 2.3** and using Perkin-Elmer GeneAmp PCR Systems 2400 and 2700. The amplified PCR products were initially ligated into PCR-BluntII-TOPO® vector (Invitrogen), before being cut and ligated into similarly digested pET15b-TEV expression vector using T4 DNA ligase. The restriction enzymes (BamHI, EcoRI and XhoI), *Pfu* DNA polymerase and T4 DNA ligase were purchased from New England Biolabs. All DNA preparations (plasmid minipreps and agarose gel extractions) were performed using Qiagen kits and protocols. DNA concentrations were measured by absorbance at 260 nm with the NanoVue™ spectrophotometer (GE Healthcare). DNA constructs were verified by DNA sequencing (DNA Sequencing Unit, University of Dundee).

Component	Volume (μL)	Step	Cycling conditions
Pfu DNA Polymerase buffer 10X	5	A. Initial denaturation	94 ^{o}C for 5 minute
dNTPs (10 mM)	1	B. Denaturation	95 ^{o}C for 1 minute
Forward primer (10 pmol/μL)	2	Annealing	44-65 ^{o}C for 1 minute
Reverse primer (10 pmol/μL)	2	Extension	72 ^{o}C for 2 minute/kbp
Template DNA	Variable (< 25 ng)	C. Final extension	72 ^{o}C for 10 minutes
Pfu DNA polymerase (1 unit/μL)	1	D. Soak	4 ^{o}C indefinite
Milli-Q water	To 50	Step B	30 cycles

Table 2.3 Standard PCR parameters and cycling conditions.

2.3 Protein expression

The recombinant proteins were expressed in *E. coli* (for strain details see section 2.1). Bacteria were transformed by mixing 15 μL competent cell culture with 1 μL plasmid DNA. The cells were left on ice for 20 minutes, before heat shocking at 42 ^{o}C for 45 seconds and cooling on ice for another 2 minutes. 100 μL of S.O.C. broth were added and the cells were incubated in a shaker at 37 ^{o}C for one hour, before plating the cells on the agar plate containing the appropriate antibiotic. For protein expression, a single colony was picked from a fresh transformation plate and incubated in 10 mL of LB supplemented with antibiotic at 37 ^{o}C overnight. The next morning, the appropriate volume of seeder culture was added to 1 L antibiotic supplemented LB to get a starting OD_{600} of 0.05. The cells were incubated at 37 ^{o}C with shaking until OD_{600} reached 0.6. Expression was induced with 0.2 mM IPTG. In brief, the expression plasmids contain the constitutive *lacI* gene encoding the repressor protein LacI that binds the lac operon engineered upstream of the gene of interest and inhibits its transcription. When IPTG is present, it binds to the repressor and displaces it from the lac operon thereby allowing T7 RNA polymerase to initiate transcription. Upon addition of IPTG, cells were grown

overnight at 22 ºC and harvested by centrifugation at 3500 *g* (Beckman J6-MC with a JA6 rotor) for 30 minutes at 4 ^{o}C. The supernatant was discarded and the bacterial pellet was resuspended in approximately 20 mL of buffer (A, C or D; **Table 2.4**), prior to the storage at -20 ^{o}C until further use.

2.4 Protein purification

Frozen cell pellets were defrosted at room temperature, followed by addition of DNase I (0.1 mg) and a single tablet of a cocktail of EDTA-free protease inhibitors (Roche). Cells were lysed using the French Press (American Instrument Company) at 16,000 psi. Insoluble debris was separated by centrifugation at 37,500 *g* (Beckman Avanti J-25 with JA25.50 rotor) for 30 min at 4 °C. The soluble fraction was filtered through a 0.2 μm Minisart® membrane (Sartorius) and loaded onto a 5 mL Ni^{2+}-charged HisTrap HP column (GE Healthcare), preequilibrated with buffer (A, C or D; **Table 2.4**). A linear imidazole concentration gradient was applied using buffer (B or E; **Table 2.4**) to elute the protein of interest. The collected fractions were analysed on polyacrylamide gels. Fractions containing the recombinant protein were pulled together and dialysed in buffer (C or F; **Table 2.4**) to remove imidazole. TEV protease, which recognises the Glu-Asn-Leu-Tyr-Phe-Gln-Gly sequence and cleaves the peptide bond between Gln and Gly, was added to remove the N-terminal histidine tag. After TEV digestion, the protein was passed through a second nickel-affinity column to remove uncleaved protein. Eluted protein was concentrated using Vivaspin and Amicon Ultra concentrators (Sartorius and Millipore) with appropriate MW cut off and then loaded onto a gel filtration column (Superdex 75 16/60 or Superdex 200 26/60; GE Healthcare) for the final purification step. The columns had previously been calibrated with molecular weight standards, thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B_{12} (1.35 kDa) (Bio-Rad). All protein purifications

were performed using ÄKTA systems (Explorer, Purifier and Prime; GE Healthcare). SDS-PAGE gel and matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (Fingerprint Proteomics Facility, University of Dundee) were used to check the final sample purity. Protein concentrations were determined spectrophotometrically using the theoretical molar extinction coefficients calculated with ProtParam (Gasteiger *et al.*, 2005).

Name	Buffer	Used for the purification of
A	50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 20 mM imidazole	SENP1
B	50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 1 M imidazole	SENP1, *Bc*CCP*
C	50 mM Tris-HCl, pH 7.5, 250 mM NaCl	SENP1, *Bc*CCP*
D	50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM imidazole	Siah1, SIP, PEG3
E	50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 M imidazole	Siah1, SIP, PEG3
F	50 mM Tris-HCl, pH 7.5, 150 mM NaCl	Siah1, SIP, PEG3

Table 2.4 Buffers for protein purification.
*Initially, *Bc*CCP protein was expressed and purified by Thomas Eadsforth (University of Dundee).

2.5 Protein structure determination by X-ray crystallography

2.5.1 Protein crystallisation theory

The crucial step in structural crystallography is to grow X-ray quality crystals. Crystallisation is a phase separation phenomenon. When the protein molecules are fully solvated, the system is at the state of equilibrium. However, as the number of molecules in solution increased, the supersaturated state is reached where there is insufficient solvent to maintain full hydration of the molecules. At this point, the molecules come out of solution to form either an amorphous precipitate or crystal nuclei thereby returning the system to thermodynamically favoured equilibrium (Weber, 1991). Precipitating agents are used to achieve protein supersaturation. The popular reagents include salts, which diminish electrostatic repulsion between proteins, and PEGs, which

compete for water molecules with proteins. Other factors affecting crystallisation include purity and concentration of macromolecule, pH, ionic strength and temperature.

2.5.2 Crystallisation techniques

The physical technique used in the study presented in this thesis to achieve supersaturation in protein solutions was vapour diffusion; in particular the hanging and sitting drop variations (**Figure 2.1**). This method relies on evaporation and diffusion of a volatile component (typically water) between solutions of different concentrations until the thermodynamic equilibrium is reached. In the initial step, the sample drops were prepared by mixing protein solution with a similar volume of precipitant solution. Generally, 1-3 μL and 0.1-0.5 μL of protein solution were used to prepare hanging and sitting drops, respectively. The drops were then suspended and sealed over a reservoir containing precipitant and/ or desiccant solution. The volumes of the reservoir solution were 1 mL and 60 μL for hanging and sitting drop, respectively. The higher concentration of precipitant in the well solution causes water molecules to travel from a drop into the reservoir until the concentration becomes equal. Over time, the precipitant and protein concentrations in a drop increase until a nucleation region is reached at supersaturated state. As the first crystals are formed the protein concentration is reduced and the system enters the metastable region, which allows crystals to grow, but does not support nucleation. Crystal growth proceeds until the protein solution becomes undersaturated.

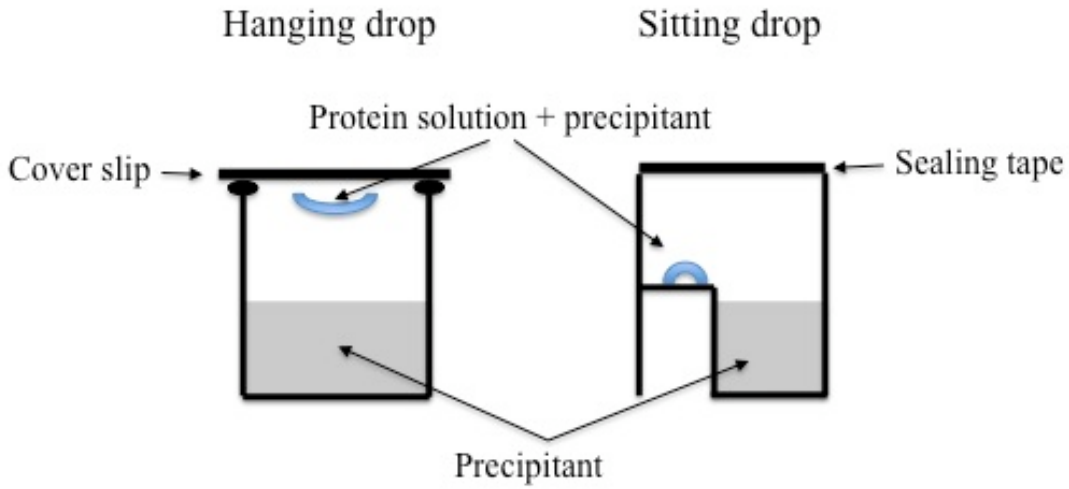

Figure 2.1 Vapour diffusion techniques.
The image shows vapour diffusion methods used in this study. Initial high-throughput screening was done using sitting drop method and the subsequent hit conditions where optimized using the hanging drop method. The blue and grey shapes represent protein crystallisation drop and reservoir solution, respectively.

2.5.3 Crystallisation strategy

It is hard to predict specific crystallisation conditions for a particular protein and numerous experiments are usually required to identify the best crystallisation conditions. However, modern robotics allows many crystallisation parameters to be screened in a short time, while consuming the minimum amounts of protein. The high-throughput Phoenix liquid handling system (Art Robbins Instruments/ Rigaku) was employed to screen several hundreds of individual crystallisation conditions known as sparse matrix screens. These were purchased from commercial sources and included Classic, PEG, MPD, $AmSO_4$ suites from Qiagen and JCSG+, PGA suites from Molecular Dimensions. Crystallisation conditions were screened in 96-well plates using a sitting drop method. The lead conditions determined from the initial screen were optimized further to produce better diffraction quality crystals. This was done manually in 24-well plates by varying different parameters in small steps and using larger volumes. Optimisation step employed a hanging drop method.

2.5.4 X-ray diffraction from crystals

Details on the theory of protein structure determination mentioned in this section can be found in Biomolecular Crystallography (Rupp, 2010). Structure determination by X-ray crystallography relies on the scattering of X-rays by the electrons found in the molecules of a crystal. The crystal acts as a three-dimensional (3D) diffraction grating, scattering the X-ray beam in many directions. However, diffracted beams called reflections of measurable intensity are only observed when the scattered waves add up in phase. In other directions the waves are out of phase and cancel one another out. Bragg's law, which treats crystals as sets of equally spaced parallel planes running in different directions, describes the conditions necessary for constructive interference to occur (**Figure 2.2**). Bragg's law, expressed as $n\lambda=2d\sin\theta$, relates the wavelength of X-ray beam (lambda, λ) to the diffraction angle (theta, θ) and the lattice spacing within a crystal (*d*). It states that waves scattered from successive planes add up constructively when the difference in the distance travelled by two equals to an integer (*n*) number of wavelengths.

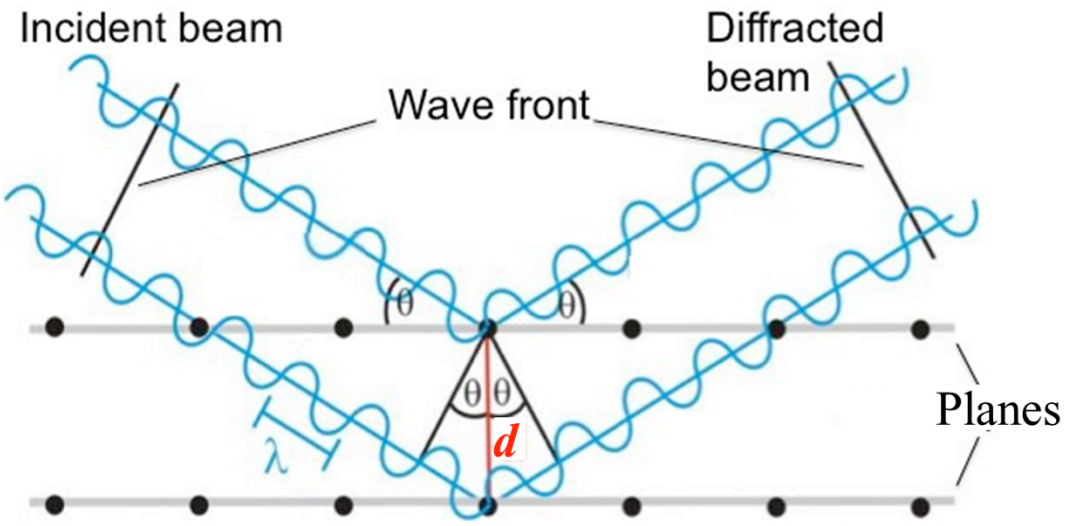

Figure 2.2 Bragg's law.
The constructive interference is observed when the path difference between two waves scattered from adjacent planes, which is $2d\sin\theta$, equals to an integer number of wavelengths $n\lambda$. Figure modified from www.microscopy.ethz.ch/bragg.htm.

2.5.5 X-ray sources

The common X-ray sources for a diffraction experiment are rotating anode generators and synchrotrons. Both were used to collect the diffraction data presented in this thesis. In an X-ray generator, the electrons emitted by the cathode are accelerated towards the anode and produce X-rays upon the impact. The material of the anode determines the wavelength of the generated X-rays. In-house generators often use the copper anode, producing CuKα radiation of λ=1.5418 Å. Filters, monochromators or X-ray mirrors are applied to get monochromatic X-ray beam focused onto the crystals. At a synchrotron facility, the X-ray radiation is generated by the beam of electrons travelling in a continuous loop with the help of the magnets. The advantage of using a synchrotron is that it produces a tunable and much more intense X-ray source than a rotating anode generator.

2.5.6 Structure determination

The goal of a crystallographic experiment is to obtain the electron density distribution, which is subsequently used to assign the atomic coordinates. The electron density map $p(xyz)$ has the following expression:

$$p(xyz) = \frac{1}{V}\sum_{hkl} \mathrm{F}\,(hkl)\exp[-2\pi\mathrm{i}(h\mathrm{x} + k\mathrm{y} + l\mathrm{z})]$$

where F(*hkl*) are the structure factors, V is the unit cell volume and *h*, *k*, *l* are the Miller indices. The electron density can be calculated from the structure factors by applying a mathematical operation called Fourier transform. Each structure factor F(*hkl*) is a complex number composed of reflection amplitude |F(*hkl*)| and its phase α(*hkl*) with the relationship:

$$\mathrm{F}(hkl) = |Fhkl|\exp(i\alpha hkl)$$

The Fourier transform requires the knowledge of both parts. The amplitudes can be easily calculated from the experimental data, as they are proportional to the square root of the measured reflection intensities. However, the phase associated with each reflection cannot be obtained directly from the diffraction pattern. This leads to the phase problem, which can be solved by a number of techniques such as multiple isomorphous replacement, anomalous scattering and molecular replacement. These methods provide approximate initial phases, whereas more accurate phases are obtained later on during model refinement. Molecular replacement was the principle technique used to solve the structures mentioned in this thesis and is described below.

2.5.7 Molecular replacement

This method uses atomic coordinates of a previously solved, structurally similar model to estimate the initial phases for the unknown protein structure. The success of molecular replacement depends on the structural homology of the proteins, which usually correlates well with their sequence identity. Molecular replacement involves the fitting of the known model structure in the correct orientation and position in the experimental unit cell. It is a generally six dimensional problem, requiring three rotation angles to orient the molecule and three parameters describing its location in the unit cell. The search method can be treated as two separate 3D problems. The Patterson function is used to determine the relative orientation of the search molecule. It is a modified electron density function, where the structure factors (amplitudes and phases) are substituted with the intensities (amplitudes squared) and is written as

$$P(uvw) = \frac{1}{V}\sum_{hkl} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)]$$

The Patterson function gives a map of the vectors between each pair of atoms in the structure and is defined by generic coordinates *u*, *v*, *w* as opposed to *x*, *y*, *z* that define

the unit cell. The advantage of the Patterson function is that it can be calculated directly using experimental data and without knowing phase angles as these are set to zero. To get the correct orientation of the search model in the unit cell, its Patterson map is calculated using the atomic coordinates and then overlaid onto the Patterson map derived for the unknown structure, which is calculated from the experimental intensities. Once the correct orientation is determined, the translation vector, positioning the search molecule in the cell unit with respect to the origin can also be obtained. The Patterson function is able to simplify the determination of rotation matrix and translation vector because its interatomic vectors belong to two categories. The intramolecular vectors are between the atoms of the same molecule, and thereby depend only on the orientation of the molecule, but not on its placement within the unit cell. These vectors are used to derive the correct rotation matrix. On the other hand, the intermolecular vectors depend on both the orientation of the molecule and its location in the cell, so these are used to obtain the translation vector. Molecular replacement is a quick method to determine the initial phases, which allow structure factors to be calculated and the subsequent construction of the electron density map.

2.5.8 Non-crystallographic symmetry averaging

The initial phases and therefore maps derived by molecular replacement can contain large errors due to the poor quality of either diffraction data or search model. One technique to improve these is to use map averaging, which is applicable when multiple copies of a protein are present within the asymmetric unit, referred to as non-crystallographic symmetry (NCS). In such instances the diffraction pattern and hence the electron density map will contain redundant information. Therefore, the electron density of the several almost identical molecules can be averaged. The repeating

features will be amplified, whereas noise and ambiguous density will be suppressed. This process can greatly improve the accuracy of the original phases.

2.5.9 Model building and refinement

Upon obtaining the first electron density map, the structure rebuilding begins. The model used in molecular replacement is changed appropriately to reflect the molecule present in the crystal. The changes to improve agreement with the electron density map include substitution, insertion and deletion of amino acids, as well as loop modifications. Following initial rebuilding, the model is refined. Refinement is an iterative process aiming to minimize the differences between the calculated and the observed structure factors. To achieve this, various parameters have to be adjusted. The parameters modified during model refinement include three positional variables (x, y, z) for each atom and the temperature factor *B*, where

$$B_j = 8\pi^2 \, U_j^2$$

U_j^2 is the mean square displacement of atom j and *B*-factor reflects atom's mobility or smearing in space. The *B*-factor model used in refinement of protein structures is usually isotropic, thereby describes only the amplitude of displacement. In contrast, higher resolution structures in small molecule crystallography allow more complicated models that describe the individual anisotropic displacement of each atom. In protein crystallography, the number of measured data is frequently not sufficient to describe all parameters defining each atom. Consequently, to obtain a reliable structure, some parameters have to be restrained towards the ideal values. The stereochemical restraints are set on the distances between atoms, torsion angles and van der Waals contacts. The ideal values for these geometrical parameters are derived from high-resolution small molecule crystallography. After each round of refinement, the new phases are computed

from the refined model. This allows a more accurate electron density map to be calculated, which is inspected and changed, and then refined again to improve it further. The cycles of manual model corrections and automated optimisation are repeated until a good correlation is achieved between calculated and experimental data.

The overall agreement is measured by crystallographic *R*-factor (R_{work}):

$$R_{\text{work}} = \frac{\sum_{hkl}|F_{\text{obs}} - F_{\text{cal}}|}{\sum_{hkl}|F_{\text{obs}}|}$$

where $|F_{obs}|$ and $|F_{cal}|$ are the observed and calculated structure factor amplitudes. However, relying on R_{work} value on its own can be dangerous, as it can lead to model over-fitting. This occurs when R_{work} is reduced artificially by refining model parameters for which no justification is present in the experimental data. For this reason, an additional R_{free} parameter is used. R_{free} is calculated using a similar equation as R_{work}, but with a fraction of randomly selected reflections (usually 5 %), which were kept separate and not included in the model refinement. R_{free} provides an independent assessment of model correctness in the refinement process. Both R_{work} and R_{free} values are reduced over the course of refinement, with refinement considered complete when R_{free} does not decrease further and when there are no significant peaks in a difference density map.

2.5.10 Software for protein structure determination and analysis

Programs from the *CCP4* (Winn *et al*., 2011) program suite were used for data processing, structure determination and refinement. Diffraction data sets were indexed and integrated with *iMOSFLM* or *MOSFLM* (Battye *et al*., 2011; Leslie, 2006) and scaled with either *SCALA* or *AIMLESS* (Evans, 2006). Determination of structures by molecular replacement was done using *Phaser* (McCoy *et al*., 2007). The output models were manually rebuilt in *Coot* (Emsley *et al*., 2010) and refined with *REFMAC5* (Murshudov *et al*., 2011). The images showing protein structures were prepared with

PyMOL (Schrödinger). Protein sequences were aligned with ClustalW2 and these alignments were illustrated using *ALINE* (Bond and Schüttelkopf, 2009). *PISA* (Protein Interfaces, Surfaces and Assemblies; Krissinel and Henrick, 2007) was used to calculate surface and dimer interface areas, while DALI (Distance matrix ALIgnment; Holm, 2010) and PDBeFold were used to perform structural alignments.

2.6 Fragment screening techniques

2.6.1 Differential scanning fluorescence (DSF)

Differential scanning fluorimetry (DSF), also referred to as the thermal shift assay, is a technique used to study protein stability in solution. The method monitors thermally induced structural changes in the protein using an environmentally sensitive fluorescence dye, which is commonly polarity sensitive. This dye's fluorescence emission is quenched in aqueous solution. However, when the protein unfolds as the temperature increases, the dye binds the hydrophobic regions and its fluorescence increases (Niesen *et al*., 2007). The point at which the concentrations of the folded and unfolded protein are equal is called the melting temperature (T_m) and is determined by a simple fitting procedure (**Figure 2.3**). This transition midpoint can be used to compare whether the protein is stabilized or destabilized under a certain condition relative to a reference. DSF is a high throughput method, which requires minimal amounts of sample. Its applications include the identification of stabilizing buffer conditions as well as ligand screening (Cummings *et al*., 2006).

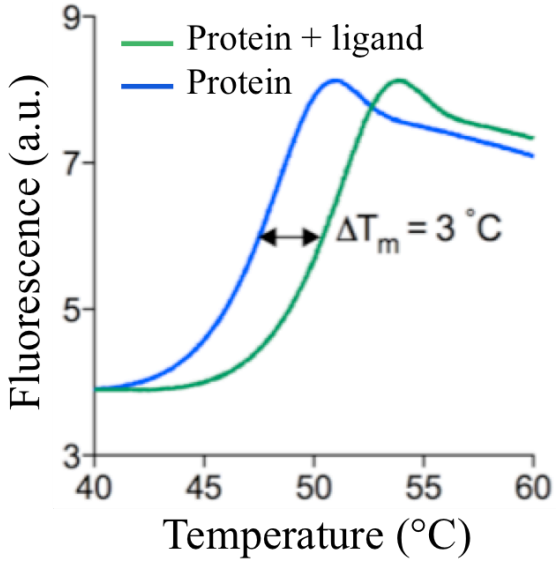

Figure 2.3 Example of the thermal profile obtained from DSF experiment.
The melting temperature of a protein was increased by 3 °C in the presence of its ligand. Figure was adapted from a paper (Scott *et al*., 2012).

DSF was used to screen SENP1 and Siah1 against the Maybridge Ro3 Fragment Library (Maybridge). The library consists of approximately 1000 compounds with physicochemical properties defined by the "Rule of Three" (Congreve *et al*., 2003). The stock compounds were prepared at 100 mM concentration in 100 % DMSO. The SYPRO Orange dye (Life Technologies) was purchased as 5000x stock concentration and was diluted 1000 times in the protein solution. The protein sample was then dispensed into the wells of a 96 well PCR plate (Thermo Scientific), followed by the addition of the fragments to give the final volume of 40 µL in each well. The final protein concentration in the screen was 10 µM. SENP1 was screened at two concentrations of fragments (1 mM and 10 mM), while Siah1 was screened against the final fragment concentration of 1 mM. The screen was run using Mx3005p qPCR instrument (Stratagene). The fluorescence intensity was monitored with excitation and emission wavelengths of 492 nm and 610 nm, respectively. Plates were scanned from 25 ^{o}C to 95 ^{o}C with a heating rate of 1 °C min^{-1}. T_m values were extrapolated fitting the raw data to Boltzmann model using Prism (GraphPad Software). Wells containing protein solution and dye were used as the reference T_m. The melting curves and thermal

shifts were visualised using a modified Microsoft Excel spreadsheet (ftp://ftp.sgc.ox.ac.uk/pub/biophysics).

2.6.2 Surface plasmon resonance (SPR)

Surface plasmon resonance (SPR) is a method for detection and quantification of biomolecular interactions (Giannetti, 2011). The SPR phenomenon occurs in an electrically conducting metal film placed between media of different refractive index. In a typical sensor chip, two media materials are the glass and the sample solution, which are separated by a thin gold film. The polarized light is focused onto the metal surface at an angle of total internal reflection and the intensity of the reflected light is being constantly monitored (**Figure 2.4**). At the certain angle of incident the reflection intensity drops as the energy of light photons is transferred into oscillating the mobile electrons of a metal plate. The angle at which the intensity decreases the greatest is known as the SPR angle. This angle is very sensitive to the refractive index of the medium adjacent to a metal surface. Therefore, changes in mass on the sensor chip surface that result from the binding or dissociation of the molecules alter the refractive index and hence the SPR angle. The changes in SPR angle are detected in real time and are plotted as response unit (RU) against time, producing a sensorgram (Biacore Sensor Surface Handbook, 2008).

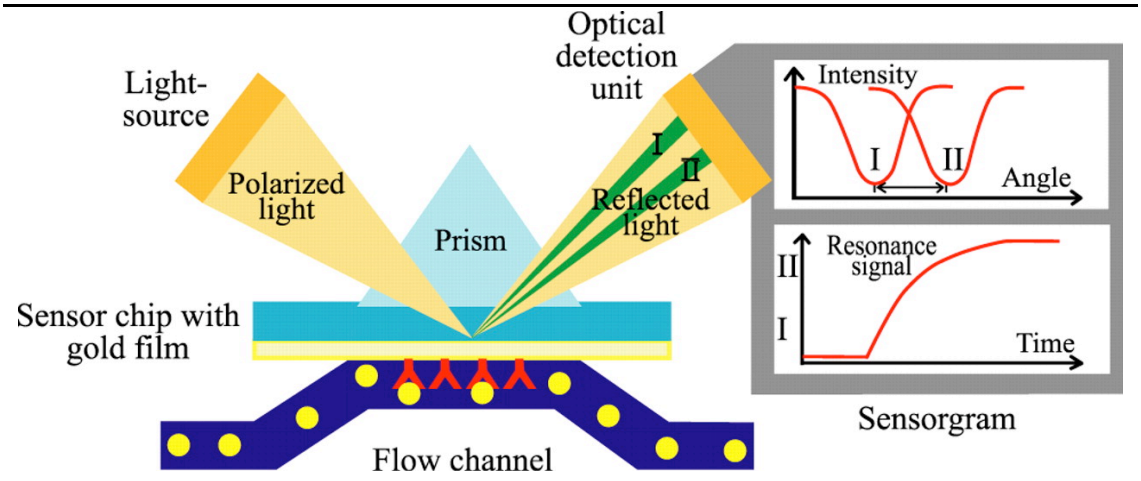

Figure 2.4 The principle of SPR.
Polarised light shines onto the sensor chip coated with a gold layer and the reflected beam is detected. At a particular angle of incidence, the light photons cause the free electrons to oscillate, resulting in the dip in the intensity. This SPR angle depends on the mass of a bound material at the sensor chip surface. The interaction between immobilised molecule (the ligand; red "Y" symbols) and its binding partner (the analyte; yellow circles) increases the overall mass on a surface, thereby increasing the refractive index and will cause the SPR angle to shift from I to II. The diagram was reproduced from Delmar *et al*., 2004.

2.6.2.1 Protein immobilisation

All SPR studies were performed at 25 °C using Biacore instruments (GE Healthcare). A fragment screen was run on a Biacore 4000; with the subsequent titration experiments run on models 3000 and 4000. Prior to the start of each new experiment, the system was cleaned using Biacore maintenance kit and a fresh NTA sensor chip was inserted, which was preconditioned and normalised according to the instructions displayed by the software. Siah1 was captured on the surface of the Sensor Chip NTA (GE Healthcare) using the N-terminal histidine tag. The NTA sensor chip's surface consists of a carboxymethylated dextran matrix with immobilised nitrilotriacetic acid (NTA), which allows the binding of histidine-tagged molecules through metal chelation such as Ni^{2+}. In addition, Siah1 was immobilised on the same chip covalently by amine coupling. The surface of the NTA sensor chip also contains carboxyl groups, which are first activated with a mixture of carbodiimide (EDC) and N-hydroxysuccinimide (NHS) to produce reactive succinimide esters. The esters react spontaneously with primary amine groups, such as free amino groups on lysine residues, to link the protein to the dextran matrix.

Following immobilisation of Siah1, Tris buffer was injected to inactivate any remaining succinimide esters on the surface. **Table 2.5** outlines the main parameters employed in immobilising Siah1 on the sensor chip's surface for fragment screening.

Sensor chip	Series S Sensor Chip NTA	
Channels	Fc1, Fc2, Fc3, and Fc4	
Spots	1,2,4,5 – protein; 3 – reference surface	
Flow rate	10 μL min^{-1}	
Siah1 concentration	200 nM	
Protein buffer	50 mM HEPES pH7.5, 150 mM NaCl, 1 mM TCEP	
Immobilisation running buffer	50 mM HEPES pH7.5, 150 mM NaCl, 1 mM TCEP, 0.005 % Tween 20	
Injection steps	Reagent	Contact time
1	500 μM $NiCl_2$	1 min
2	EDC/NHS	7 min
3	Protein sample	10-15 min
4	100 mM Tris pH 8.0	7 min

Table 2.5 Immobilisation conditions for Siah1

2.6.2.2 Fragment library screening

The SPR screen was performed using AstraZeneca's in-house fragment library, which contained 3072 compounds. The molecular weight (MW) of the fragment library ranged from 130 Da to 466 Da. The average MW was 228 Da, equating to 17 non-hydrogen (heavy) atoms. 2 μL of each fragment at 25 mM concentration in 100 % DMSO were dispensed into 384 well plates by the in-house Compound-Storage Unit. 98 μL of the sample preparation buffer were added later using the automatic dispensing robot to give a final concentration of 500 μM. The plates were then sealed and vortexed in a centrifuge to get solutions to the bottom of the wells and remove the air bubbles. The main solutions used to set up the fragment screen are presented in **Table 2.6**.

Running buffer	50 mM HEPES pH7.5, 150 mM NaCl, 1 mM TCEP, 0.005 % Tween 20, 2 % DMSO
Sample preparation buffer	50 mM HEPES pH7.5, 150 mM NaCl, 1 mM TCEP, 0.005 % Tween 20
Positive control	Stock solution of 100 mM plectin peptide was diluted to 2.5 μM in the running buffer
Negative control	Running buffer
Solvent correction samples	Sample preparation buffer containing 1.4 %, 1.6 %, 1.8 %, 2.0 %, 2.2 %, 2.4 %, 2.6 % and 2.8 % DMSO
Preparation of compounds	Stock solutions at 25 mM in 100 % DMSO were diluted to 500 μM in sample preparation buffer

Table 2.6 Solutions used in a fragment screen of Siah1

Positive and negative control samples were injected every 50 cycles to monitor protein activity. Solvent corrections were performed every 100 cycles to check the influence of DMSO. The flow rate was 30 μL min^{-1}, with injection and wash steps set at 60 seconds each.

2.6.2.3 Fragment hit characterisation

Titration experiments were performed to confirm and characterise the hit compounds. Six concentration points, varying from 2 μM to 2 mM, were chosen to cover the expected range of fragment binding affinities. 100 mM stock solutions were prepared by dissolving solid compounds in 100 % DMSO. These were subsequently diluted to 2 mM in sample preparation buffer. The 2 mM samples were then used to perform four fold dilutions in the running buffer in five increments, resulting in 2 mM, 500 μM, 125 μM, 32 μM, 8 μM and 2 μM solutions. The later SPR screen of the near neighbours of the top five hits obtained from a fragment screen was performed at the same six concentrations. However, these compounds were prepared in 100 % DMSO at the appropriate concentrations directly by the in-house Compound Storage Unit, and the sample preparation buffer was then added to reach the final titration concentrations. Solvent correction samples, positive and negative controls were prepared and run using

the same method employed in the case of the fragment screen. The flow rate and injection time were left the same, with the wash step increased to 240 seconds. SPR data were analysed using Biacore 4000 Evaluation Software and Scrubber2 (BioNavis).

2.6.3 Nuclear magnetic resonance (NMR)

Nuclear magnetic resonance (NMR) spectroscopy is an important technique for structure determination and ligand screening. Detection of ligand binding can be studied by two main types of NMR experiment. They are known as ligand and target based methods depending on which part of the protein-ligand complex is being observed. Ligand based 1D methods detect changes in the ligand spectrum, occurring when the energy is transferred to the ligand from either the protein or solvent upon its binding to the protein (Lepre, 2011). The widely used 1D NMR techniques are saturation transfer difference, water ligand observed via gradient spectroscopy, and T_1 and T_2 relaxation experiments (Mayer and Meyer, 1999; Dalvit *et al*., 2000; Hajduk *et al*., 1997). The target based 2D methods, such as ^{1}H-^{15}N HSQC experiments that require ^{15}N-labelled protein, are used to screen and validate compound binding simultaneously (Shuker *et al*, 1996). The ^{15}N-^{1}H HSQC spectrum shows the peaks corresponding to each amide nitrogen in the protein, thereby revealing individual protein residues (except proline). The position of the peaks is sensitive to the chemical environment and the chemical shifts are observed upon ligand binding. This approach can provide direct information about the localisation of ligand binding, as the largest perturbations are expected to occur at the ligand binding site. The advantages of 1D NMR over 2D methods include no upper limit on MW of the proteins, no need for isotopically labelled protein and they require less protein (Scott *et al*., 2012; Hubbard, 2011; Holdgate *et al*., 2010). However, 2D methods tend to be more robust with low false positive rates and can provide the binding site and more accurate K_d measurements.

The NMR experiments were performed with the help of Dr. Vajpai Navratna (AstraZeneca). The primary screen of SENP1 with 384 fragments from AstraZeneca's in-house fragment library was performed using 1D NMR experiments. The stock compounds were prepared in DMSO-d6 at a 100 mM concentration. 1 μL of each compound in the set of six was subsequently added to 500 μL of the protein solution (20 mM Tris-HCl pH8.0, 50 mM NaCl, 1 mM TCEP, 5 % D_2O) containing SENP1 at 10 μM concentration. Therefore, the final volume of sample in the NMR tubes was 506 μL and the final concentration of fragments was approximately 200 μM. The runs containing the mixtures of compounds in a buffer were used as a reference. The 2D target observed ^{1}H-^{15}N HSQC experiments were performed with isotopically labelled SENP1 and Siah1 proteins for further characterisation of individual compounds. These were run at a protein concentration of 100 μM and the compounds were then added in steps, ranging from the initial to the final concentrations of around 50 μM to 4 mM.

Chapter 3

Results and discussion: Seven-in-absentia homolog 1

3.1 Aims

The aims of this project were to investigate human Siah1 by producing soluble protein and performing a fragment-based screen against it. Further objectives were to determine the crystal structure of Siah1 and use X-ray crystallography to validate potential hits. Siah1 is a scaffolding protein, where the interface with its substrates is formed via a flat shallow surface, making it a difficult target for drug discovery. This study would inform as to whether or not a fragment-based approach can identify the ligands of Siah1 and which screening technique might be applicable.

3.2 Recombinant Siah1 expression and purification

The gene coding for Siah1 without the RING domain was obtained from a commercial source and was subsequently inserted into a pET15b-TEV vector for protein expression (see section 2.2.1). The protein was expressed in Bl21 (DE) Gold cells and purified using nickel affinity and gel filtration chromatography as described in sections 2.3 and 2.4, respectively. ^{15}N-labelled Siah1 for NMR studies was expressed in the appropriate minimal media and purified using a similar protocol. However, isotopically enriched Siah1 and Siah1 used for SPR studies were purified with the histidine tag left on by skipping the TEV digestion step. The protein used for DSF screening and for structural studies was purified without the tag. Irrespectively, Siah1 eluted from the gel filtration column as a symmetric peak. Siah1 with the tag removed eluted with a mass of approximately 39 kDa (**Figure 3.1**). The theoretical mass of Siah1 is 21,897 Da, which suggests Siah1 is a homodimer in solution. This is consistent with the earlier observations reporting Siah1 as a dimeric protein (Polekhina *et al*, 2002). The purity of the sample during purification was monitored by SDS-PAGE (**Figure 3.1**). Typical final yields were 10 – 15 mg L^{-1} of cell culture. The purified protein was kept at 4 °C for short-term storage and flash frozen in liquid nitrogen for long-term storage at -80 °C.

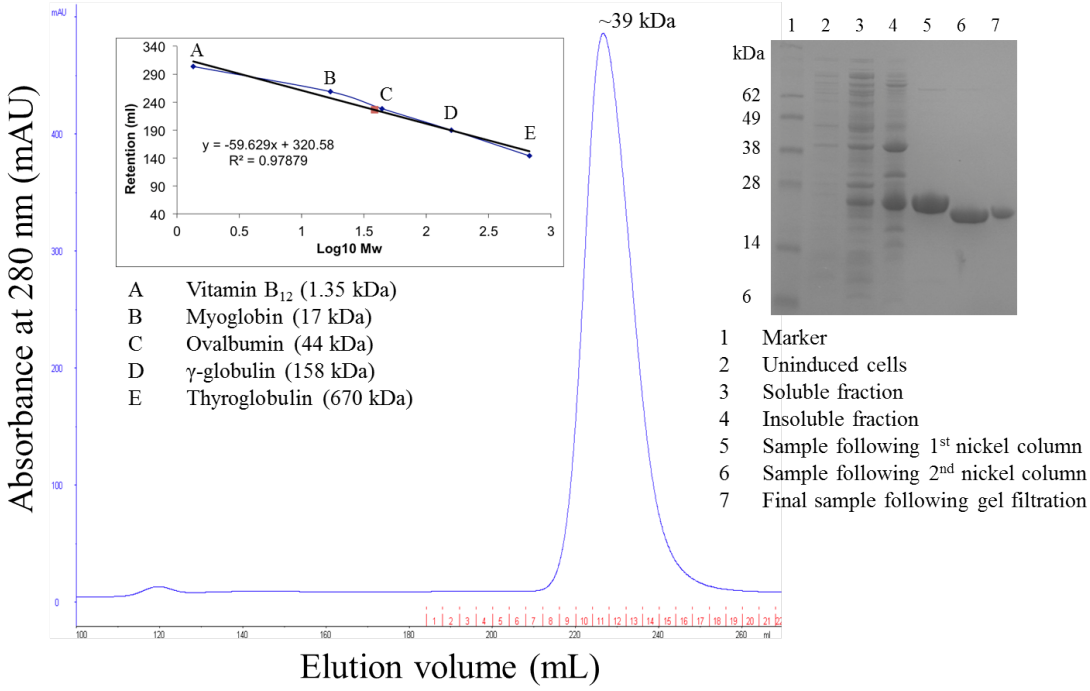

Figure 3.1 Size exclusion chromatography and SDS-PAGE analysis.
The column was equilibrated using five standards labelled A to E. The red dot on the calibration curve marks the eluted peak. The peak corresponded to approximately 39 kDa, equivalent to a Siah1 dimer. SDS-PAGE analysis after each purification step is shown on the right hand side.

3.3 Fragment screening

3.3.1 DSF

The thermal profile was measured to check that Siah1 is applicable to DSF prior to beginning the fragment screen. A single melting curve was observed, suggesting the presence of a globular domain with a melting temperature of 64.5 ºC (**Figure 3.2**). The fragment screen was performed as described in section 2.6.1. The final concentrations of protein and fragments were 10 μM and 1 mM, respectively. The results are presented in **Figure 3.3**. Thermal shifts larger than 2 ^{o}C were considered to be significant. This threshold value, corresponding to about three times the standard deviation of the measured melting point, was chosen based on the previous DSF studies in the laboratory. Two compounds produced temperature shifts above this set value. However,

they had untypical melting curves with decreased fluorescence levels, suggesting non-specific binding and protein denaturation, and were not investigated further.

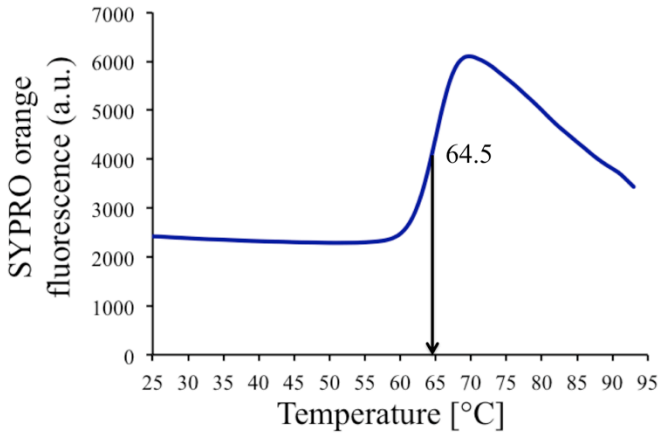

Figure 3.2 The thermal profile of Siah1. Protein is in 50 mM Tris-HCl pH7.5, 50 mM NaCl buffer and displays single melting curve with T_m of 64.5 ºC.

Menadione was reported to inhibit Siah1 activity in the Meso Scale electro-chemiluminescent assay and in *in vivo* experiments using 293T cells in the previous study (Shah *et al*., 2009). Its binding to Siah1 was tested using DSF at 1 mM concentration, but no changes in the T_m were observed.

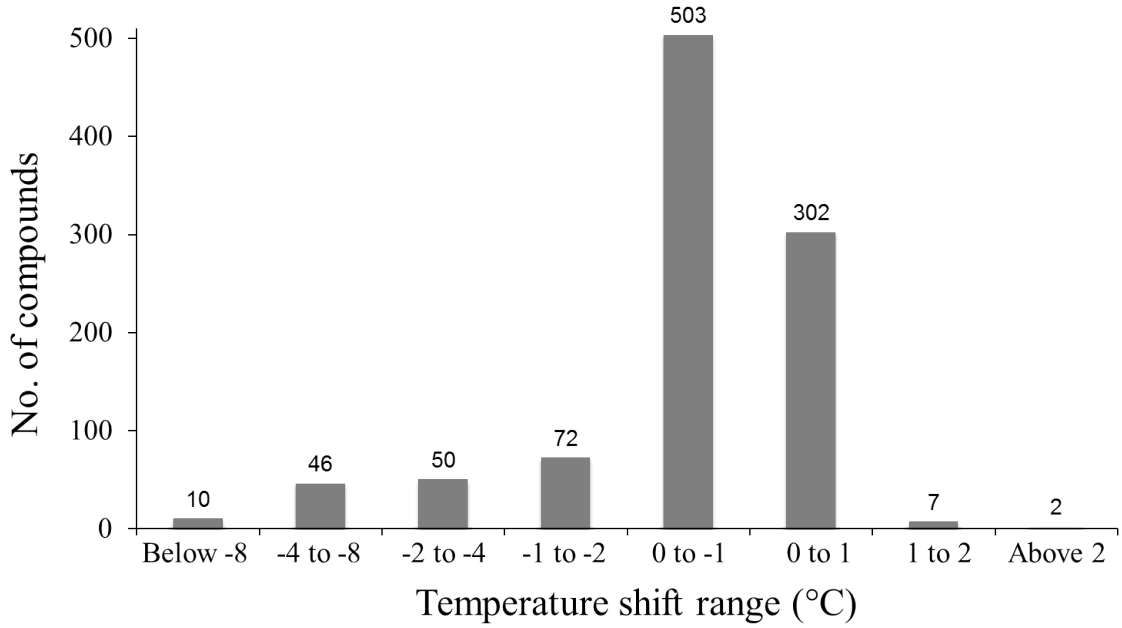

Figure 3.3 Results from the Maybridge library screen using DSF.
Each column represents the number of compounds that produced a temperature shift within a particular range.

3.3.2 SPR

Siah1 was immobilised on the sensor chip's surface using the histidine tag as described in section 2.6.2.1. The immobilisation levels of Siah1 were approximately 3500 RU. A search of the literature for a potent Siah1 binder that could be used as a positive control revealed the plectin-1$^{95\text{-}117}$ peptide. It had a K_d value of around 100 nM as measured by SPR (House *et al*., 2003). This 23 residue and a shorter nine-residue peptide were purchased and tested (see **Table 2.1** for amino acid sequences). **Figure 3.4** shows the sensorgrams and a dose response curve for the plectin peptide injections at 10 different concentrations ranging from 19.5 nM and going up two fold to 10 μM. The data were fitted using a Langmuir binding isotherm

$$R_{obs} = \frac{R_{max} \times [C]}{K_d + [C]}$$

in which [C] is the concentration of injected ligand, R_{obs} is the observed ligand binding and R_{max} is the maximum binding capacity, when all protein binding sites are occupied. It is the simplest binding model and makes a number of assumptions: the ligand is monovalent, the protein on the surface is homogeneous and all binding events are independent. The Langmuir binding model yielded the K_d values of about 28 μM and 225 nM for the shorter and a longer plectin peptide, respectively, and was applied to fit the data from compound titration experiments.

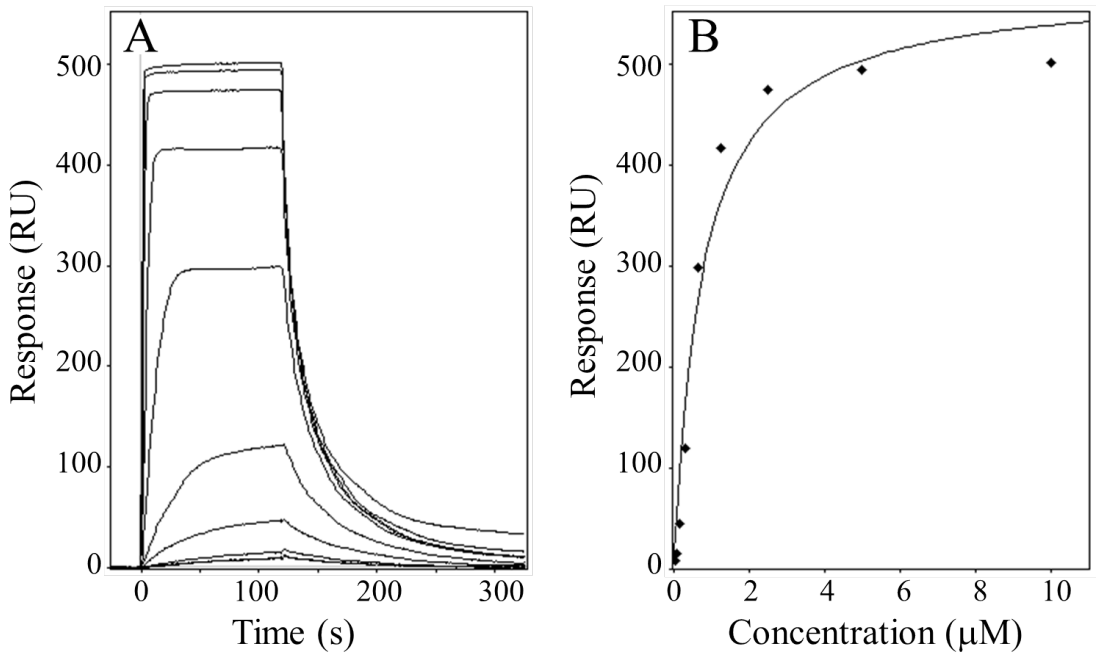

Figure 3.4 SPR data showing the binding of the positive control to Siah1.
Sensorgrams (A) and dose-response curve (B) are presented. The plectin peptide was injected at the following concentrations: 19.5 nM, 39 nM, 78 nM, 156 nM, 312 nM, 625 nM, 1.25 μM, 2.5 μM, 5 μM and 10 μM. Siah1 immobilisation levels were 5000 RU, resulting in R_{max} of 500 RU. The peptide saturated the protein's surface at concentrations above 2.5 μM, thereby this concentration of the positive control was chosen for the screen.

The plectin peptide was subsequently used as a positive control to measure the functional integrity of immobilised Siah1. Protein activity levels were determined with the assumption that plectin peptide binds Siah1 at 1:1 stoichiometry using the following equation

$$\text{Protein activity level} = \frac{\text{protein MW}}{\text{ligand MW}} \times \frac{\boldsymbol{R_{max}}\ (\text{RU})}{\text{immobilised protein level (RU)}}.$$

R_{max} was measured experimentally to be 350 RU and using this value the initial activity level of immobilised Siah1 was calculated to be approximately 90 %. The plectin peptide was injected at even intervals during the fragment screen and revealed a gradual decrease in surface activity over the course of the run. Sample responses were later adjusted to compensate for this drop in surface activity. DMSO solvent corrections were

applied to eliminate the DMSO contributions to the sample responses. Fragment data were also referenced for blank injections and blank reference flow cell.

The R_{max}, corresponding to the upper limit of responses for 1:1 fragment protein binding was calculated from the rearranged protein activity equation. The R_{max} of 30 RU was determined using the average MW of all fragments. The lower cutoff level was subsequently set at 10 RU. This value was derived from the Langmuir equation based on 500 μM fragment screening concentration and the K_d values set at 1 mM and higher. 3072 fragments were screened in total and the screening results are presented in **Figure 3.5**. The two horizontal lines show the upper and lower cutoff levels. The spots represent individual fragments and are coloured according to their response levels. The red spots belong to the fragments whose responses exceeded 1:1 stoichiometry, suggesting they either bind to multiple sites or aggregate non-specifically on the surface of the protein.

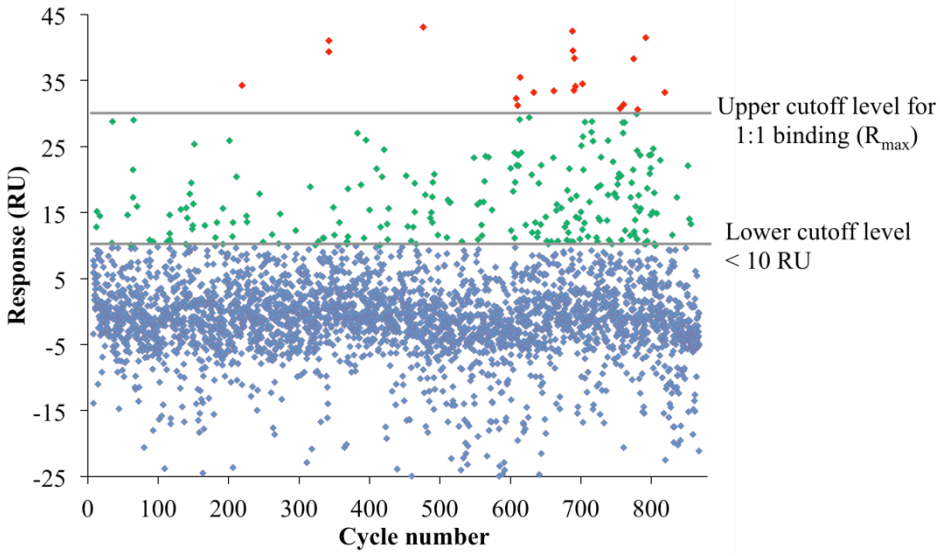

Figure 3.5 A plot of the responses obtained from the fragment screen.
Response levels for the compounds are plotted against injection cycle numbers. The upper and lower cutoff levels are shown by the horizontal lines. The blue spots indicate fragments whose response levels were below the lower cutoff of 10 RU. Fragments whose responses fell within the range of 1:1 binding are shown as green spots. The red spots represent fragments whose stoichiometry was above 1:1.

39 compounds that exhibited responses above the lower threshold were selected for further investigation, following the manual inspection of their sensorgrams. These were subsequently screened at six concentrations in the range between 2 μM and 2 mM. The responses and sensorgrams obtained were used to classify 39 fragments into four groups (**Figure 3.6**).

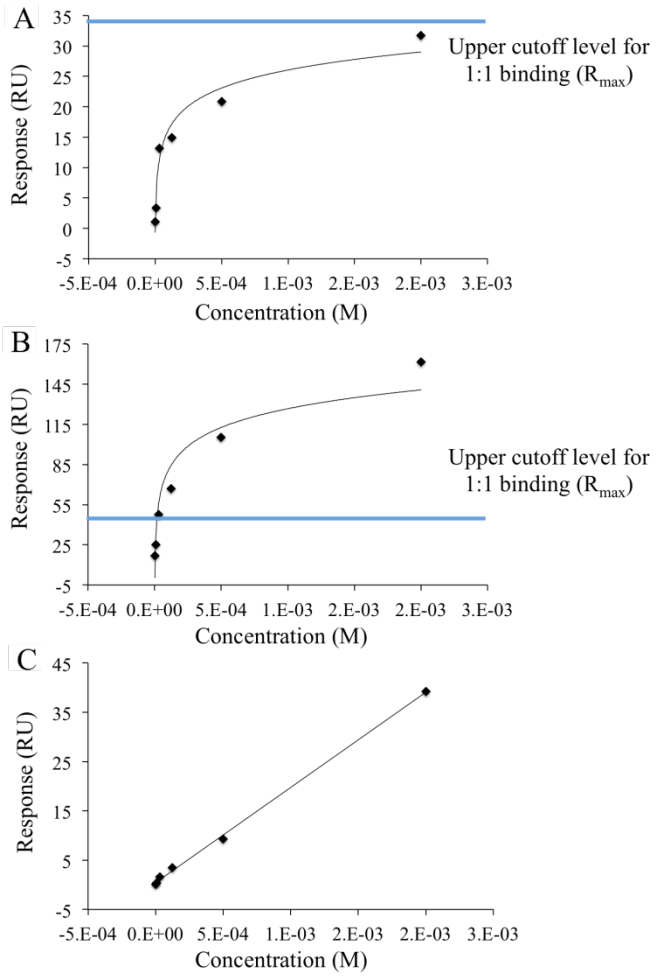

Figure 3.6 Compounds exhibit different types of binding behaviour.
These were: typical transient binding (A), promiscuous binding (B) and concentration-dependent aggregation (C).

Five fragments exhibited typical transient binding behaviour. These compounds bound reversibly, eventually dissociating to baseline and either obeyed 1:1 binding model or reached saturation levels above R_{max}, but remained below 2:1 stoichiometry. Seven fragments were classified as promiscuous binders. They showed nonstoichiometric

binding, where binding responses exceeded R_{max} by more than a factor of two and some bound to the protein surface irreversibly. The third group contained 12 fragments that showed concentration-dependent aggregation. Compounds belonging to this group behaved well at low concentrations, where they showed no signs of promiscuous or nonstoichiometric binding. However, their responses became significantly disproportionate from the levels expected for a 1:1 binding model at higher concentrations. The final group included 15 compounds whose sensorgrams failed to show binding responses above the lower threshold level of 10 RU.

The search of AstraZeneca's compound library was performed to identify the compounds structurally similar to five fragments that showed transient binding behaviour. 146 near neighbour compounds were selected for the dose response experiments. These were performed at six concentrations as described in section 2.6.2.3. The data were analysed in the same way as the data from the titration experiments with 39 fragment hits from the initial fragment screen. 21 compounds were chosen for further structural studies using X-ray crystallography (section 3.6). **Figure 3.7** depicts the chemical structures and SPR data of three of those 21 fragments. The top fragments exhibited a K_d of around 1 mM. However, the majority of fragments failed to reach saturation levels at the highest screening concentration and their data were not reliable enough for a kinetics analysis.

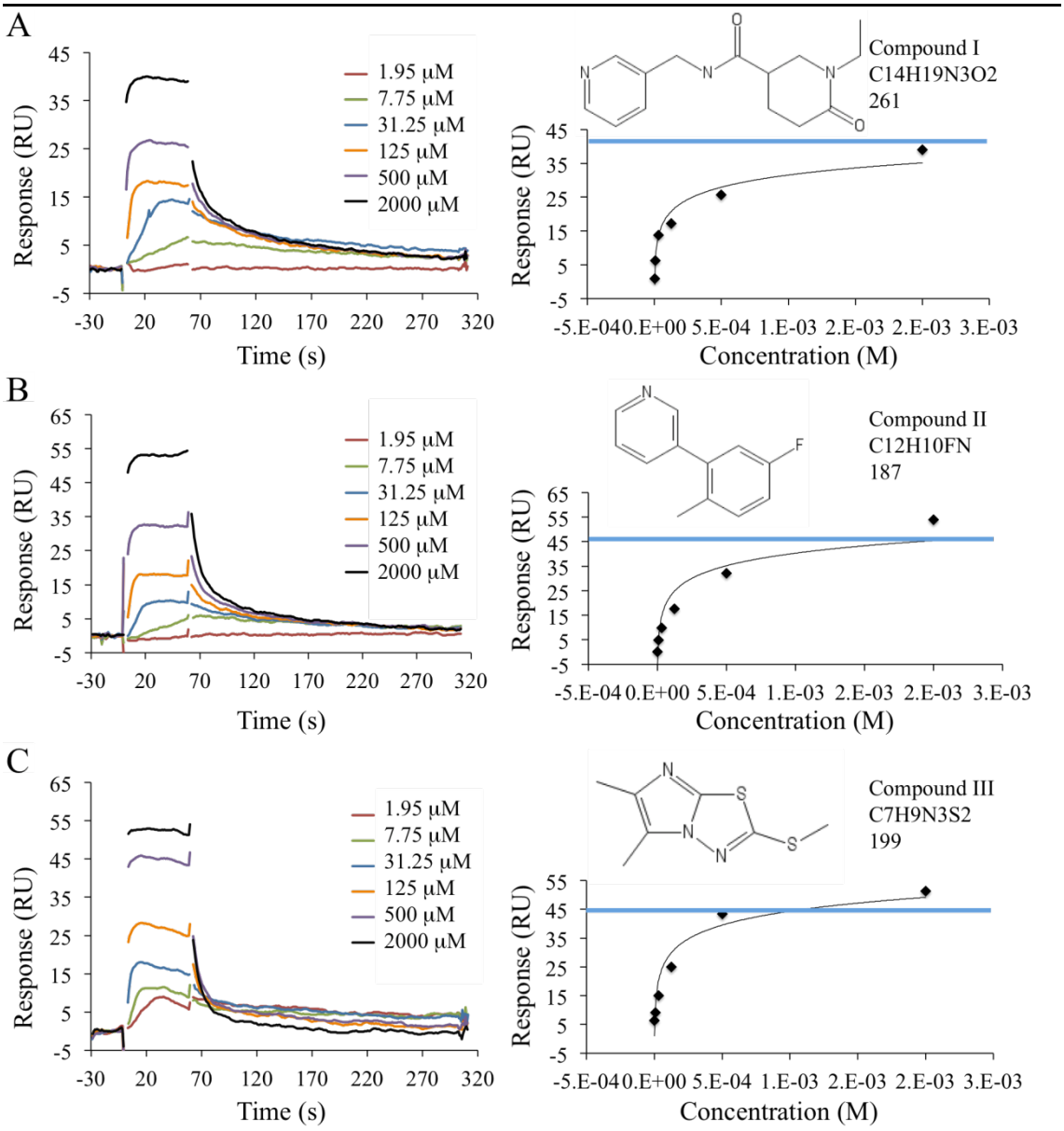

Figure 3.7 Examples of the potential fragment hits identified.
Sensorgram data from the titration experiments are shown on the left-hand side. Dose-response curves and chemical structures are presented on the right-hand side. Chemical formulae and molecular weights are shown next to the structures. Blue lines indicate the upper cutoff levels for 1:1 binding (R_{max}). The compounds were classified as typical transient binders. Compound I (A) exhibited 1:1 binding behaviour. Compounds II (B) and III (C) had responses exceeding the R_{max} expected for 1:1 interaction.

3.3.3 NMR

NMR experiments were run to test the binding of two synthetic peptides and menadione, which was reported to inhibit Siah1 function in the literature (Shah *et al.*, 2009). These were performed as described in section 2.6.3. The binding of the nine and 23-residue peptide was observed using 1D and 2D NMR experiments, respectively. The

measured K_d for the nine-residue peptide was 40 μM. The K_d for the plectin peptide was not determined. The ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled Siah1 were measured prior and following the addition of menadione. The overlaid spectra revealed changes in peak position indicating binding (**Figure 3.8**). However, protein precipitation was observed during these experiments at menadione concentrations above 1 mM. Thus, NMR failed to conclusively determine whether the menadione interaction is specific or non-specific.

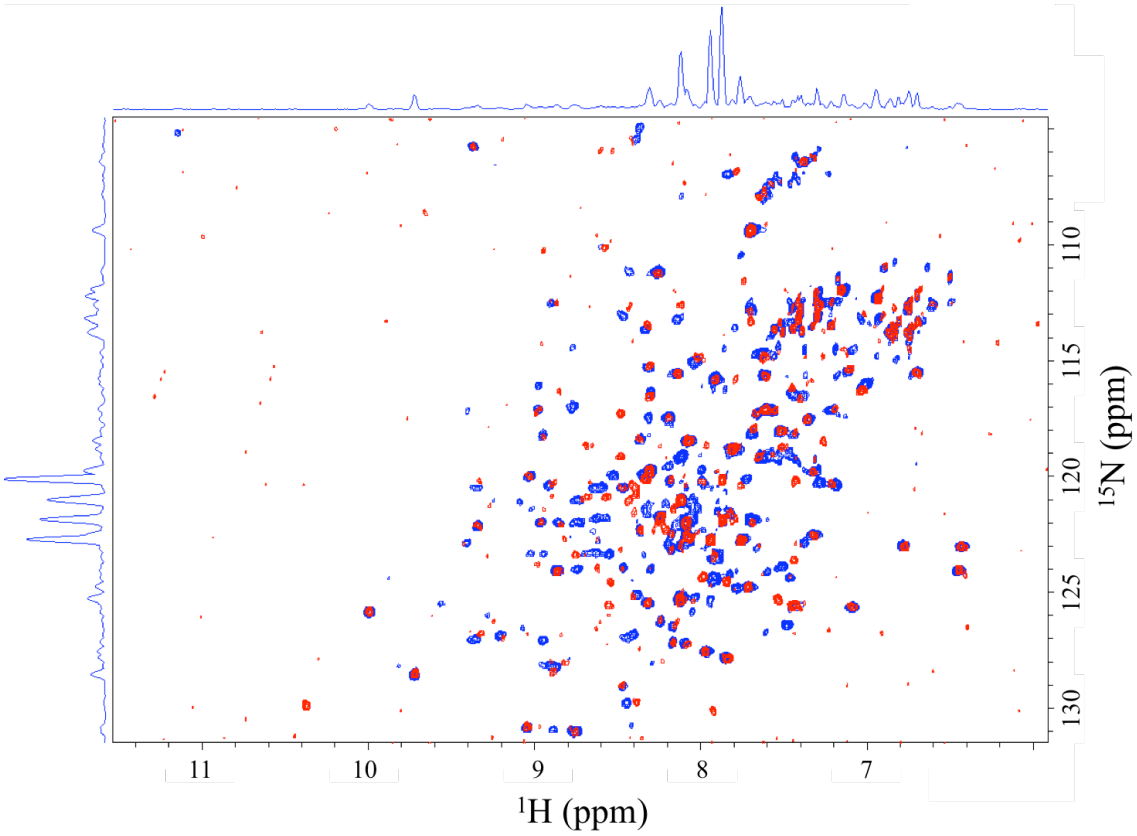

Figure 3.8 Superposition of 2D ^{1}H-^{15}N HSQC spectra of Siah1 and Siah1-plectin peptide.
Overlaid spectra show Siah1 (red) and Siah1-plectin peptide complex (blue) in 50 mM HEPES pH7.5, 100 mM NaCl and 5 % D_2O.

3.4 Crystallisation and data collection

The crystal structures of Siah1 in complex with the fragments identified in section 3.3.2 were required to validate their binding. Thus, conditions were sought-after that produce apo-crystals of Siah1 for subsequent ligand soaking experiments as well as conditions promoting co-crystallisation of the protein-ligand complexes. Initial crystallisation

screens were performed using a sitting drop vapour diffusion method with commercially available formulations (see section 2.5.3). Crystallisation occurred within three days at a protein concentration of 10 mg mL^{-1} in several conditions. Two conditions were refined to yield diffraction quality crystals. Optimised crystals grew from equal volumes of protein solution (15 mg mL^{-1} in 15 mM Tris-HCl pH7.5, 30 mM NaCl and 10 mM DTT) and reservoir solution containing either 100 mM MES pH6.5, 1.5 M $MgSO_4$ (**Figure 3.9**) or 100 mM HEPES pH7.0, 1.45 M Li_2SO_4. Crystals had similar shape and size irrespective of the reservoir used. Crystals were soaked in a cryoprotectant containing the mother liquor solution supplemented with 20 % (v/v) glycerol before being cryocooled at -173 ºC. Diffraction data were collected in-house using a Rigaku MicroMax-007 rotating-anode X-ray generator (copper K_α, λ = 1.5418 Å) coupled to an R-AXIS IV^{++} image plate detector. The protein crystallised in space group *I*222 with unit cell lengths *a* = 75.14 Å, *b* = 104.59 Å, c = 133.16 Å (**Table 3.1**).

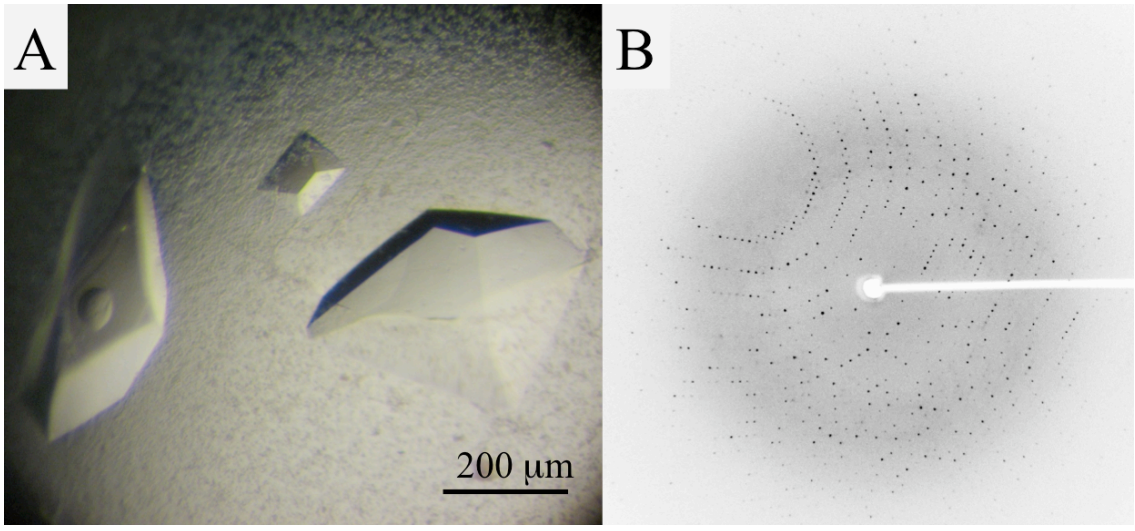

Figure 3.9 Siah1 crystals.
(A) The optimised crystals grew over three days in the reservoir condition 100 mM MES pH6.5, 1.5 M $MgSO_4$. (B) The crystals diffracted in-house to 1.95 Å.

3.4.1 Structure solution and refinement

The structure was determined by molecular replacement with *Phaser* (McCoy *et al*.,

2007) using a molecule of human SENP1 (PDB code 2a25; Santelli *et al*., 2005) as a search model. The R_{work} and R_{free} factors of the output solution subjected to a round of rigid body refinement using *REFMAC5* (Murshudov *et al*., 2011) were 41.5 % and 42 %, respectively. Model building was done in *Coot* (Emsley *et al*., 2010) and the structural refinement performed using *REFMAC5*. Tight local non-crystallographic symmetry restraints were employed in the initial rounds of refinement and TLS parameters (Translation/ Libration/ Screw) were applied in the later stages. 30-35 residues at the N-terminal end and three internal regions, which were absent in the search model, were modelled in once the electron density was clear. The final model includes residues 91-282 and 95-282 of the native Siah1 sequence in chains A and B, respectively (**Table 3.1**). Chain A also contains an additional Met residue at the N-terminus left from the histidine tag. In addition, the model contains water molecules and a number of different ligands (Tris, sulphate, glycerol, Cl^- and Zn^{2+}).

	Siah1
Space group	*I*222
Unit cell dimensions: *a*, *b*, *c* (Å)	75.14, 104.59, 133.16
Resolution[a] (Å)	44.99 - 1.95 (2.06 - 1.95)
No. of reflections	141688 (11394)
Unique reflections	38340 (5357)
Completeness (%)	99.3 (96.0)
Multiplicity	3.7 (2.1)
$\langle I/\sigma I \rangle$	19.3 (4.5)
Wilson *B* ($Å^2$)	29.2
Residues	
Chain A	91-282
Chain B	95-282
Water / Cl^- / sulphate / Tris / glycerol / Zn^{2+}	218 / 4 / 3 / 2 / 3 / 4
R_{merge}[b] (%)	3.7 (17.8)
R_{work}[c], R_{free}[d] (%)	19.8 / 23.9
Mean *B*-factors ($Å^2$)	
Protein	17.7
Ligands and waters	29.4
Ramachandran plot[e] (%)	
Most favoured	97.4
Additional allowed	2.3
Outliers	0.3
R.m.s.d. on ideal values[f]	
Bond length (Å) / bond angles (°)	0.01 / 1.30

Table 3.1 Crystallographic statistics of Siah1.
a. Values in parentheses refer to the highest resolution shell. **b.** $R_{\text{merge}} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. **c.** $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. **d.** R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from refinement calculations. **e.** (Chen *et al*., 2010). **f.** (Engh and Huber, 1991).

3.5 The Siah1 structure

3.5.1 Structure quality

The structure was determined at 1.95 Å resolution with R_{work} and R_{free} values of 19.8% and 23.9%, respectively. The asymmetric unit consists of two polypeptide chains with the solvent content of 59 % and V_m (Matthews, 1968) of 2.99 $Å^3$ Da^{-1}. The polypeptides are arranged as a dimer consistent with the dimer observation on the size exclusion column and in other published structures of Siah1 (House *et al*., 2003; Polekhina *et al*., 2002; Santelli *et al*., 2005). Continuous, well-defined electron density is observed for

most of the structure; apart from a few residues at the N-terminus of Chain B. Ramachandran analysis reveals that 99.7 % of the amino acids are located in the allowed region. A comparison of the determined Siah1 structure (subunit A) and that of PDB entry 2a25 gives an rmsd of 0.61 Å for 140 Cα atoms.

3.5.2 Overall structure

Siah1 is a dimeric protein, which forms a Z-shaped structure (**Figure 3.10**). Each monomer contains two zinc finger domains and a C-terminal substrate binding domain (SBD), which forms the dimerization interface with an equivalent domain of the second subunit. Superposition of two full-length monomers gives a high rmsd value of 5.36 Å for 181 Cα atoms. This is due to the shift in the position of the N-terminal zinc finger domains relative to each other and is likely imposed by the crystal packing. The result is that the Zn^{2+} ion in this domain is displaced by about 14.5 Å when the monomers are overlaid. Superposition of the monomers without their N-terminal zinc fingers gives a lower rmsd of 0.886 Å for 150 Cα atoms.

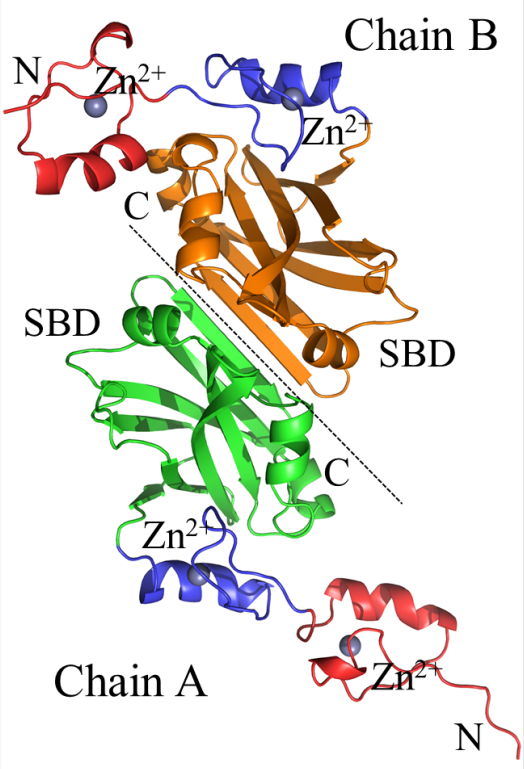

Figure 3.10 Overall structure of Siah1.
The homodimer is shown as a cartoon. The zinc fingers are coloured red and blue, the substrate binding domain (SBD) of monomer A and B are coloured green and orange, respectively. The N- and C- termini are labelled accordingly. The Zn^{2+} ions are shown as grey spheres.

The SBD forms two antiparallel β-sheets of four strands each placed on top of one another (**Figure 3.11**). The first contains strands β2, β3, β4 and β7, while the second is made of strands β1, β8, β5 and β6. A further strand β0 is situated parallel to β2, connecting the SBD with an adjacent zinc finger. In addition, the SBD contains three short α-helices (α1-α3) that link β4 with β5, β6 with β7 and β7 with β8.

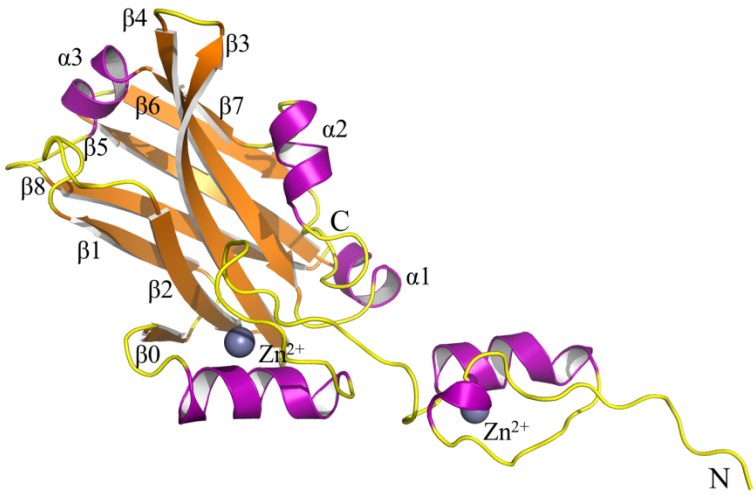

Figure 3.11 Structure of the Siah1 monomer.
Ribbon diagram is showing the SBD and two zinc finger domains. The α-helices and β-strands are coloured purple and orange, respectively. The N- and C-terminal positions are labelled, as are the α-helices and β-strands. The Zn^{2+} ions are shown as grey spheres.

The total surface area of a monomer is about 10800 $Å^2$ of which 1100 $Å^2$ are buried upon dimer formation. Therefore, approximately 10 % of the surface area of each monomer is involved in dimerisation. The Siah1 homodimer is held together by a network of hydrogen bonds, salt-bridge interactions and van der Waals forces. The majority of these intermolecular contacts occur between the residues found on strands β6 of each monomer. For example, Arg232 and Arg233 on β6 of one subunit make salt-bridge interactions to Asp255 and Glu237 of the partner subunit (data not shown). The equivalent arginine residues in the partner subunit make similar contacts to the residues in the first subunit; thereby eight salt-bridges are formed in total. A further six hydrogen bonds are formed between the main chains of these two antiparallel β6 strands. In addition, the interactions between residues at the dimer interface are mediated by a number of water molecules. This extensive network of different intermolecular

interactions suggests a tight association of the two monomers.

3.6 Crystallisation screening

Crystallisation was performed with compounds that exhibited binding by SPR to validate their interaction and to obtain information about the binding sites (see section 3.3.2). In total, five fragments identified in the initial fragment screen and 21 near-neighbours were used in co-crystallisation and crystal soaking experiments. Compounds were used in co-crystallisation at 5 mM and 10 mM concentrations, with 5 % and 10 % DMSO, respectively. Over ten X-ray data sets were collected and analysed, but compounds could not be seen in the electron density structure. In crystal soaking experiments, the apo Siah1 crystals were transferred into the solutions containing individual compounds at 20 mM concentration (20 % DMSO) before being tested. The majority of these crystals formed cracks and diffracted below 4.0 Å, which is likely due to the high concentration of DMSO. However, two data sets were collected, but upon structure determination the electron density map did not show the presence of the compounds. Therefore, none of the compounds identified by SPR could be validated as true hits. In addition, crystallisation was performed with the inhibitory plectin peptide at 1-10 mM concentrations. No crystals were obtained, suggesting peptide binding disrupts the crystallisation interface under these conditions.

3.7 Comparison of fragment screening methods

DSF was not able to generate any potential hits for Siah1 in this study. The technique does not appear to be sensitive enough to detect the weak binding affinities, which are usually associated with fragments. Similar results have been obtained from fragment screening studies with other protein targets in the laboratory (data not shown); thereby providing further support for this observation. The exception was observed for

Pseudomonas aeruginosa penicillin binding protein 3, with thermal shifts of over 15 ºC and 5 ºC in the presence of known antibiotics and a number of fragment hits, respectively (Sainsbury *et al*., 2011). External studies reported that the hit rate obtained by DSF is generally a number of times lower than with NMR or SPR (Hubbard and Murray, 2011). Together these studies show that DSF is still applicable for certain protein targets, and while it is unable to reliably identify weak binders, it can be used for the identification of high affinity compounds. DSF is one of the quickest and cheapest methods for compound screening, not requiring protein labelling or prior knowledge of its function, and should still be considered as a preliminary screening method.

The initial SPR screen identified 39 hits for Siah1, suggesting the technique is more applicable for fragment screening. The screen was performed at 500 μM, which was higher than the usual fragment concentration used in the SPR screen for targets like protein kinases. The fragment concentration chosen represented a compromise between the lower hit rate observed for protein-protein interaction targets and increased nonspecific fragment interactions seen at high screening concentrations (Hämäläinen *et al*., 2008). Although the screen generated 39 hits, the subsequent titration experiments showed that some of them were false positives from non-specific binders and aggregates. The false positives could have been rapidly identified by immobilising a Siah1 unrelated protein on one of the spots of the sensor chip to be used as a negative control during the screen. SPR is similar to DSF in that it allows a high throughput format sufficient for screening libraries containing thousands of fragments within days and it requires even smaller quantities of protein. However, sufficient levels of the protein should be immobilised to achieve a high signal to noise ratio to reliably detect the binding of low MW fragments. SPR is also highly sensitive and slight variations in buffer composition could interfere significantly with the data interpretation. Unlike

DSF, SPR can provide information on affinity and kinetics, which are determined from the time dependent fragment association-dissociation response. Studies indicate that SPR is able to detect the binding of fragments in the affinity range of 0.1-1000 μM (Hubbard and Murray, 2011). The detection of lower affinities is limited predominantly by aqueous solubility of the fragments being tested. In this study, the fragment hits were characterised at the highest concentration of 2 mM. This limited the reliable determination of their affinities to 1 mM. Most of the identified compounds had a lower K_d as they failed to reach protein saturation at 2 mM.

The Maybridge Ro3 library (Maybridge) was screened with DSF, while the SPR screen was performed using an SPR tailored library (AstraZeneca) assembled from the in-house compounds and from vendor catalogues. Both libraries are intended for screening a diverse range of targets, but only the Maybridge library complied with the rule of three (Congreve *et al*., 2003; see section 1.3.1). The size of its fragments varied between 93-298 Da, with an average of 178 Da (www.maybridge.com). In comparison, the SPR library contained fragments from 130 Da to 466 Da, with a mean of 228 Da. Neither library was screened with the second technique; therefore the direct comparison of the two cannot be made. However, it is likely that the hit rate with the SPR when screening Maybridge library would have been lower. The technique is not sensitive enough to robustly detect the binding of the fragments found at the lower end of MW in the Maybridge library; thereby the response levels for the fragments of less than 100 Da would be too low to fit the data accurately.

3.8 Discussion

Ubiquitination is a post-translational modification controlling many cellular processes, including the regulation of protein degradation together with 26S proteasome (Krämer *et al*., 2013). In this modification, ubiquitin is attached to the target protein, with the

process specificity defined by the E3 ubiquitin ligases such as Siah1. Siah1 functions in Ras, DNA-damage, and hypoxia signalling pathways (House *et al*., 2009). Studies showed its inhibition impairs tumour growth and metastasis, making Siah1 an attractive therapeutic target in cancer (Wong and Möller, 2013).

This work presents the results obtained from fragment screening experiments, which employed two different approaches. The first screen was run with DSF and did not identify any potential hits for Siah1. This and other experiences of screening with DSF, suggest it is not suitable as a frontline hit generation method, although it can still be useful as a preliminary screen for some targets. The initial SPR screen identified 39 hits out of 3072 fragments tested, thereby giving the hit rate of 1.3 %. However, more than half of those were later confirmed to be false positives. Thus, the final hit rate is close to 0.5 %, which was a value reported by Vernalis in its screens of about 1200 fragments against other protein-protein interaction targets (Roughley and Hubbard, 2011). The results demonstrate that SPR is sensitive enough technique for identifying the weak binding fragments for protein-protein interaction targets, supporting previous observations (Nordström *et al*., 2008; Hämäläinen *et al*., 2008).

Co-crystallisation and crystal soaking experiments were performed in an attempt to validate SPR hits as true binders and determine their binding mode. Thus far, the data sets collected did not show the presence of ligands. Alternative soaking methods could be tried, such as transferring the apo-crystals gradually to a series of fragment solutions of increasing concentration, as well as extending the soaking period. Other biophysical methods such as isothermal titration calorimetry (ITC) and NMR could be explored to verify the hits. ITC works by measuring the heat released upon the interaction during a series of titrations of ligand into a protein solution and yields full thermodynamic parameters like enthalpy and entropy (Scott *et al*., 2012). Two disadvantages of ITC are

that it requires large quantities of protein and the ligands tested need to be very soluble. Ligand-observed and protein-observed NMR can be used to measure affinity, with the later also providing information about the binding site. The drawback is the need for high amounts of isotopically labelled protein. Biochemical assays can often be used to validate hits, but no straightforward assay is available for Siah1.

The plectin peptide was shown to bind Siah1 with a K_d of around 100 nM and to inhibit its activity (House *et al*., 2003). The same peptide was tested in this study to be used as a positive control in an SPR screen. A K_d of approximately 225 nM was determined, which is not far off the reported value. The initial value was measured using biosensor-based method with the peptide immobilised on the sensor surface, while SPR experiments performed here used immobilised protein and the peptide was injected in aqueous solution. In addition, two experiments were run under different buffer conditions, which might explain the slight variations in the determined K_d values. The binding of plectin peptide was also detected with protein-observed NMR (data not shown). The interaction site of such peptides containing a conserved VxP motif in their sequences with Siah1 is known (**Figure 1.5**; House *et al*., 2006). Therefore, SPR competition assays can be performed in future to test indirectly whether the compounds identified in the fragment screen bind at the same site as the plectin peptide.

The apo-structure of Siah1 was determined at an improved resolution of 1.95 Å (see section 3.5). The structure contains the residues in three loop regions that were missing in a search model as well as the whole zinc finger motif adjacent to the RING domain. This more accurate model of Siah1 may aid future structural experiments investigating Siah1-ligand complexes.

Chapter 4

Results and discussion: Sentrin specific protease 1

4.1 Aims

The objectives of this project were to probe the druggability of the catalytic domain of human cysteine protease SENP1 using a fragment-based approach. This study would inform whether fragment screening is applicable for generating the starting chemical entities for further drug development process for such targets. In addition, the crystal structure of SENP1 was sought in order to follow up the hits obtained from a fragment screen with structural studies.

4.2 Recombinant SENP1 expression and purification

The pHISTEV30a plasmid containing the gene encoding the catalytic domain of SENP1 was a kind gift of Ron Hay (University of Dundee). The protein was expressed and purified as described in sections 2.2 and 2.4. ^{15}N-labelled SENP1 for NMR studies was expressed in the minimal media and purified using the same protocol. SENP1 eluted from the gel filtration column as a single peak with a mass of around 21 kDa. A monomer has the theoretical mass of 28,042 Da (**Figure 4.1**). The purity of the final sample was estimated at >95 % as checked by SDS-PAGE (**Figure 4.1**). A final yield was approximately 30 mg L^{-1} of cell culture. The purified protein was stored at 4 °C and flash frozen for storage at -80 °C until further use.

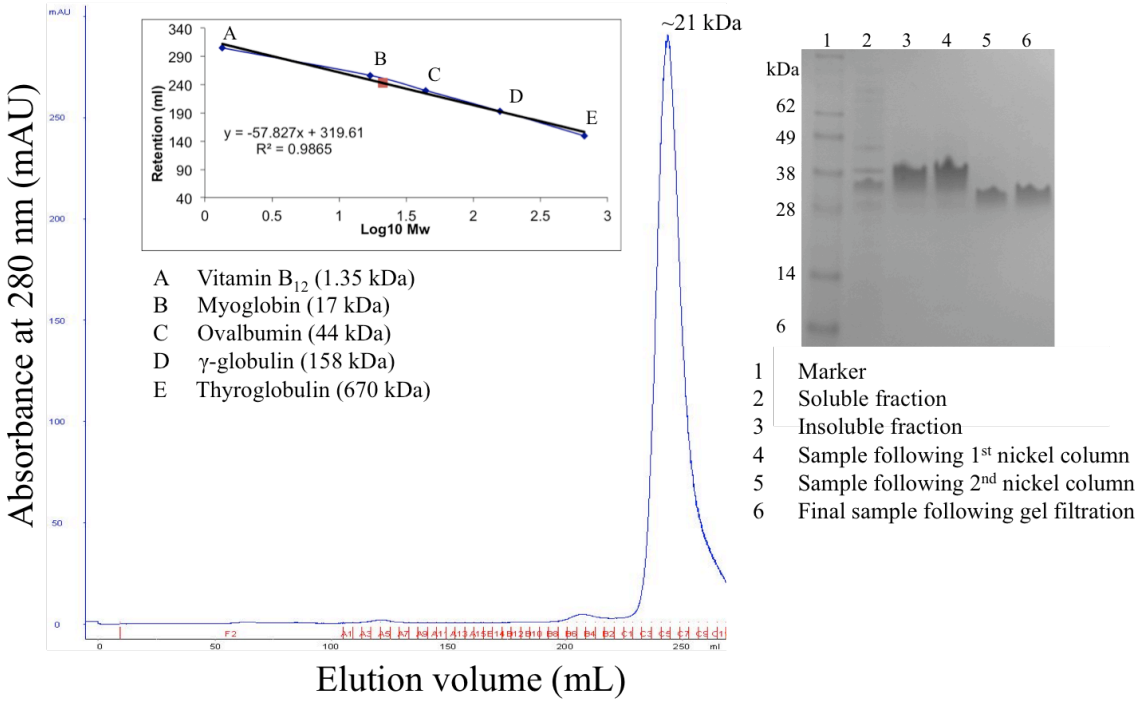

Figure 4.1 Size exclusion chromatography and SDS-PAGE analysis.
The calibration curve was plotted using five standards labelled A to E. The eluted peak is represented by a red dot on a calibration curve. The peak corresponded to approximately 21 kDa; suggesting SENP1 is a monomeric protein. SDS-PAGE gel, showing the protein sample after each purification step, is depicted on the left.

4.3 Fragment screening

4.3.1 DSF

A DSF run of the protein in a buffer was performed first to establish if SENP1 is amenable to this technique. The thermal profile of SENP1 showed a single melting curve, indicative of a folded protein. The melting temperature was 44.8 ºC in 20 mM Tris-HCl pH8.0, 50 mM NaCl buffer used in a fragment screen (**Figure 4.2**).

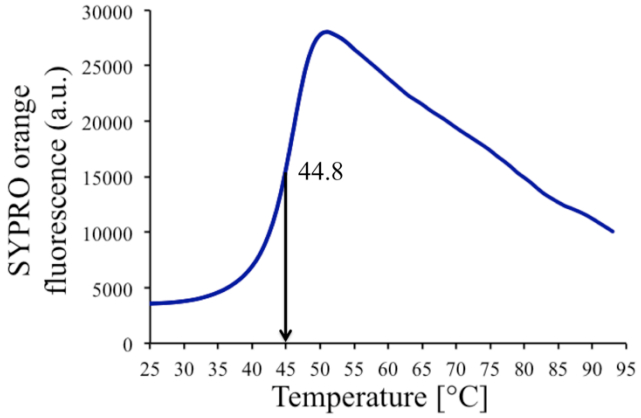

Figure 4.2 The thermal profile of SENP1.
Protein displays a single melting curve with T_m of 44.8 ºC.

The fragment screen was performed as outlined in section 2.6.1. The compounds were initially screened at the final concentration of 1 mM. Compounds that stabilised the protein by over 2 ^{o}C were classified as hits. This arbitrary cut-off value was set based on the previous experience of compound screening in the laboratory. None of the fragments produced temperature shifts above the threshold, with the best compounds stabilizing SENP1 by 1 ^{o}C (**Figure 4.3**). The original screen was repeated at the fragment concentration of 10 mM. This meant screening SENP1 at a higher DMSO concentration because the stock compounds were prepared in 100 % DMSO. The effect of DMSO on SENP1 had to be tested first as DMSO generally destabilises the protein. SENP1 was screened at 10 different DMSO concentrations ranging from 2.5 % to 25 % (**Figure 4.3**). The results showed SENP1 was still stable at the DMSO concentration of 20 %, thereby allowing the fragment library to be screened at 10 mM concentration. The results are shown in **Figure 4.4**. 27 compounds had temperature shifts above the 2 ^{o}C mark. These were the same compounds that gave the biggest temperature shifts in the initial 1mM screen; thereby there was a good correlation between the temperature shifts observed and the compound concentration used. However, the requirement to use 10 mM concentration, suggest the compounds have low affinity for SENP1 and they were excluded from further characterisation.

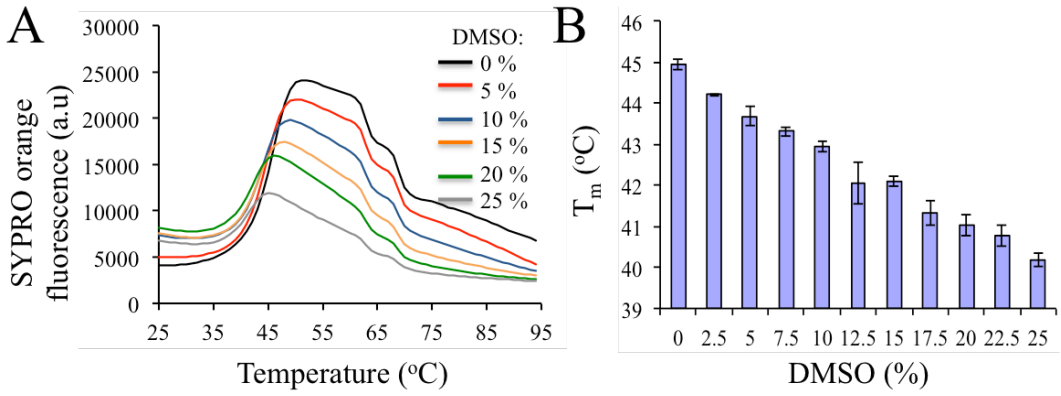

Figure 4.3 The effects of DMSO on SENP1 stability.
(A) The thermal profiles of SENP1 at different DMSO concentrations. (B) The melting temperatures of the SENP1 unfolding transition decreased with increased DMSO concentration. Values shown represent the mean T_m measured in quadruplicate ± SD.

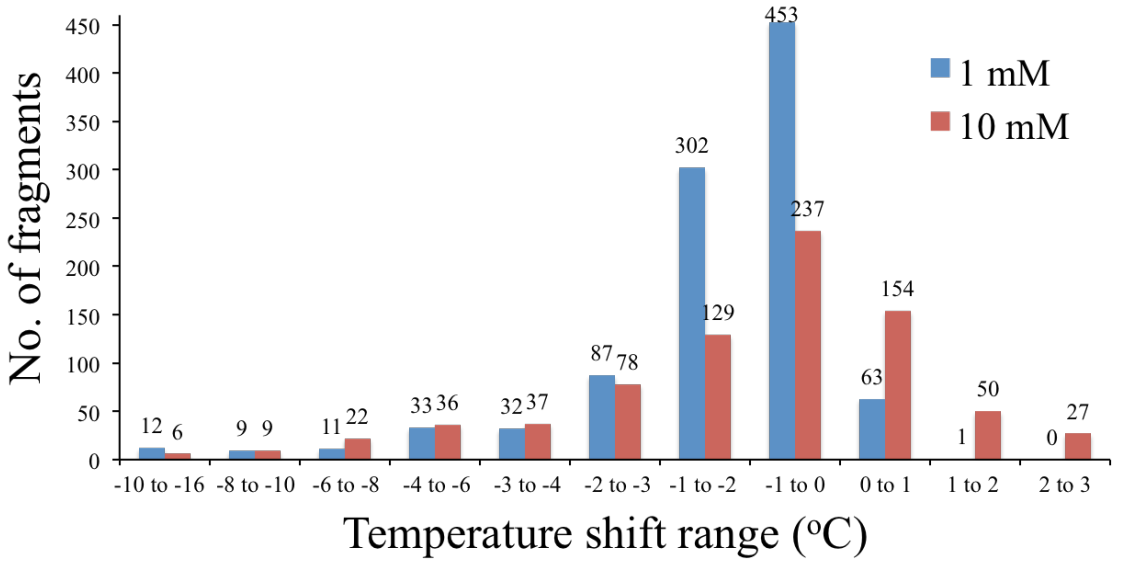

Figure 4.4 Results from the Maybridge library screen using DSF.
Fragments were screened at 1 mM (blue) and 10 mM (red) concentrations. Each column represents the number of compounds that produced a temperature shift within a particular range.

4.3.2 NMR

When screening compound libraries using 1D NMR, it is preferable to have a tool compound, which is known to bind the target as this greatly speeds up the interpretation of the results. For this purpose, a short synthetic peptide corresponding to the last four C-terminal residues of SUMO (Gln-Thr-Gly-Gly) was purchased from GenScript and tested for SENP1 binding. In addition, the binding of benzbromarone, which was reported to inhibit SENP1 in a fluorescence resonance energy transfer based assay (Martin, 2008), and its three analogues was tested. 1D NMR methods could not detect the peptide binding and the binding of benzbromarone analogues was weak with K_d above 1 mM. Due to the absence of a tool control, the screen of 384 compounds from a library of 1100 fragments was initiated without a positive control as described in section 2.6.3. K_d values for the top 11 fragments from this screen were estimated to be 10 mM. The fragment screen was subsequently terminated due to the absence of a tool compound, which could be used in validation of potential hits.

4.4 Crystallisation and data collection

Initial crystallisation screens were set up as sitting drop vapour diffusion experiments in a 96-well plate format (see section 2.5.3). Crystals grew at a protein concentration of 10 mg mL^{-1} and in the reservoir formulation 100 mM MES pH6.5, 1.8 M $(NH_4)_2SO_4$ and 10 mM $CoCl_2$. This was further refined using hanging drop vapour diffusion method in 24 well plates. Diffraction quality crystals appeared after two days from equal volumes of reservoir solution (1.8 M $(NH_4)_2SO_4$, 50 mM $CoCl_2$ and 100 mM MES pH 6.5) and protein solution (20 mg mL^{-1} in 20 mM Tris-HCl pH 8.0, and 50 mM NaCl; **Figure 4.5**). Crystals were cryo-protected in reservoir buffer containing 20% glycerol. Crystals were initially characterised in-house using a Rigaku MicroMax-007 rotating-anode X-ray generator coupled to an R-AXIS IV^{++} image plate detector. Diffraction data were then collected at the European Synchrotron Radiation Facility beamline ID29 (λ = 0.977 Å). The data were indexed with *MOSFLM* (Leslie, 2006) and scaled using *SCALA* (Evans, 2006) from the *CCP4* program suite (Winn *et al.*, 2011).

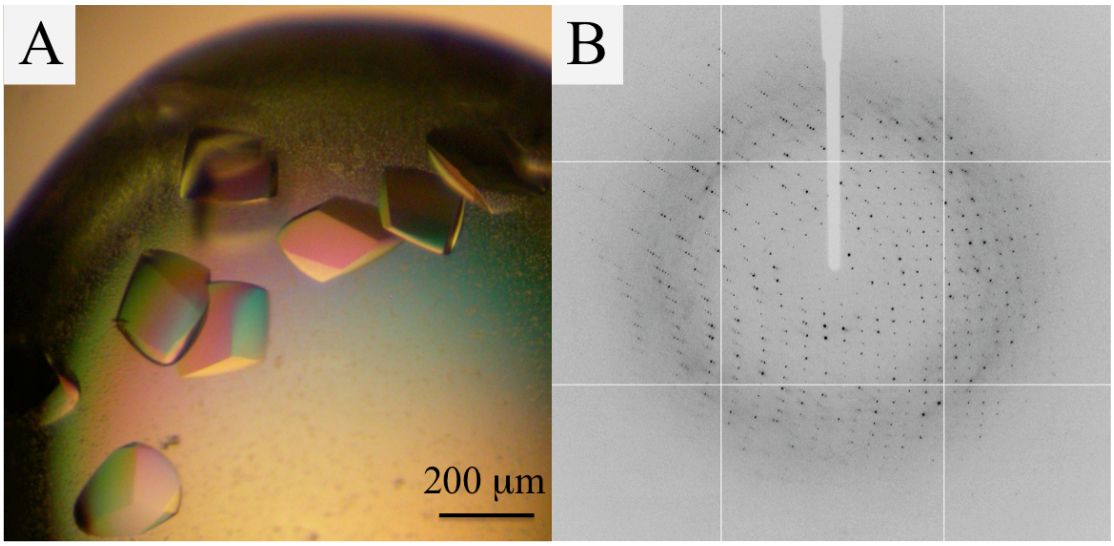

Figure 4.5 SENP1 crystals.
(A) The optimised crystals grew after two days. (B) The crystals diffracted at the ESRF synchrotron to 2.40 Å.

4.4.1 Structure solution and refinement

The structure was solved by molecular replacement with *Phaser* (McCoy *et al*., 2007) using the coordinates of a monomer (chain A) of human SENP1 (PDB code 2yic; Shen *et al*., 2006) as a search model. The model was manipulated using *Coot* (Emsley *et al*., 2010) and refinement carried out using *REFMAC5* (Murshudov *et al*., 2011). Strict non-crystallographic symmetry restraints were applied in the early stages of the refinement. Water and glycerol molecules were included in the model and a positive peak (>10σ) in the F_o-F_c difference map was assigned as a Co^{2+} ion. The refinement continued until no significant changes in R_{work} and R_{free} were observed and when inspection of the difference density map suggested that no further corrections or additions were required. The stereochemistry of the structure was checked using *MolProbity* (Chen *et al*., 2010). Data collection and structure refinement statistics are shown in **Table 1**.

	SENP1
Space group	$P3_121$
Unit cell dimensions: *a, b, c* (Å)	71.17, 71.17, 199.99
Resolution[a] (Å)	45 - 2.40 (2.53 - 2.40)
No. of reflections	120621 (17892)
Unique reflections	23808 (3415)
Completeness (%)	99.8 (100.0)
Multiplicity	5.1 (5.2)
$<I/\sigma I>$	8.5 (2.5)
Wilson B ($Å^2$)	59.507
Residues	
Chain A	418-644
Chain B	419-644
Water / glycerol / Co^{2+}	70 / 6 / 1
R_{merge}[b] (%)	9.0 (42.8)
R_{work}[c], R_{free}[d] (%)	23.1 / 31.3
Mean *B*-factors ($Å^2$)	
Protein	59.8
Ligands and waters	36.4
Ramachandran plot[e] (%)	
Most favoured	91.6
Additional allowed	8.0
Outliers	0.4
R.m.s.d. on ideal values[f]	
Bond lengths (Å) / angles (º)	0.02 / 1.62

Table 4.1 Crystallographic statistics of SENP1.
a. Values in parentheses refer to the highest resolution shell. **b.** $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - <I(hkl)>|/ \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $<I(hkl)>$ is the mean value of $I_i(hkl)$ for all *i* measurements. **c.** $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. **d.** R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from refinement calculations. **e.** (Chen *et al*., 2010). **f.** (Engh and Huber, 1991).

4.5 The SENP1 structure

4.5.1 Structure quality

The crystal structure of SENP1 was determined at 2.4 Å resolution with R_{work} and R_{free} values of 23.1% and 31.3%, respectively. The discrepancy of 8.2 % might be due to the fact that optimum restraints were not selected during refinement. Thus, it might be useful to test different weighting schemes, like *B*-factor restraints, during refinement. The asymmetric unit contains two polypeptides, referred to as A and B, with an estimated solvent content of 53 % and a V_m (Matthews, 1968) of 2.61 $Å^3$ Da^{-1}. The two

subunits located within the asymmetric unit superimpose closely with an rmsd of 0.72 Å for 215 Cα atoms as calculated using *LSQKAB* (Kabsch, 1976). In addition, the asymmetric unit consists of 70 waters, six glycerol molecules and one Co^{2+} ion, contributing to intermolecular interactions relevant to crystallisation of the enzyme. The ion is tetrahedrally coordinated by Glu430 and His640 from one molecule and the corresponding residues from a symmetry related molecule (**Figure 4.6**). A Ramachandran plot reveals that 99.6 % of residues are found in the allowed region. Most of the amino acids have a well-defined electron density, apart from the side chains of a few lysine residues. These atoms were included in the model, but their occupancy was set at zero. There are no major conformational differences between human SENP1 structure presented here and that of PDB entry 2iyc, with an rmsd of 0.43 Å for 222 Cα atoms when subunits A are superimposed. However, the Co^{2+} ion was inadvertently omitted in the earlier structure of this enzyme, which was crystallised in similar conditions containing $CoCl_2$. The presence of the Co^{2+} ion was confirmed by re-refining entry 2iyc (data not shown).

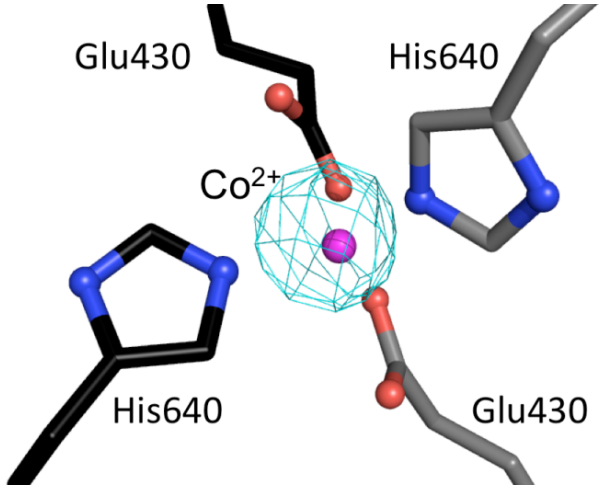

Figure 4.6 The F_o-F_c omit difference density for Co^{2+}.
F_o are observed, F_c the calculated structure factors. The map is contoured at 5σ (cyan chicken wire). Glu430 and His640 from two asymmetric units are represented as a ball-and-stick model. Carbon positions are coloured grey for molecule B and black for the symmetry related molecule A. Oxygen and nitrogen positions are coloured red and blue, respectively; the purple sphere represents Co^{2+}.

4.5.2 Overall structure

SENP1 has an overall protein fold that places it within the cysteine protease superfamily. The structure of the catalytic domain of SENP1 can be subdivided into two

subdomains. The N-terminal segment is formed by two antiparallel β-strands (β1 and β2) and five α-helices (α1, α2, α3, α7 and α8) (**Figure 4.7**). The C-terminal subdomain consists of a five stranded mixed β-sheet (β3-β7), where the middle strand β5 is antiparallel to the other four, and this sheet is surrounded by three α-helices (α4-α6). The SENP1 active site resembles the active sites of other cysteine proteases like yeast Ulp1 and human SENP2, in that it contains a conserved catalytic triad of cysteine, histidine and aspartate (Mossessova and Lima, 2000; Reverter and Lima, 2004; Li and Hochstrasser, 1999). In SENP1, the nucleophilic Cys603 is coordinated by the general base His533, which is in turn stabilised by Asp550. The Cys603 is located at the N-terminus of the central helix α7, while His533 and Asp550 are situated at the front of β5 and at the end of β6, respectively (**Figure 4.7**). The N- and C-terminal ends are located close together, but far away from the active site.

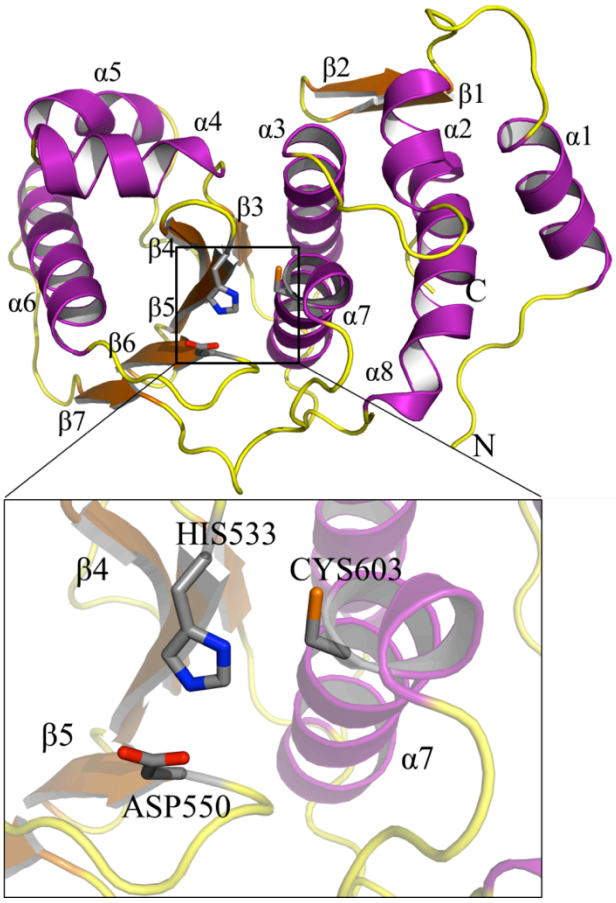

Figure 4.7 The structure of SENP1.
A ribbon diagram showing α-helices and β-strands coloured purple and orange, respectively. The N- and C-terminal positions are labelled, as are the α-helices and β-strands. The catalytic site of SENP1 is enlarged. The conserved catalytic triad of CYS603, HIS533 and ASP550 making up the active site are displayed as ball-and-stick models. Atomic positions are coloured: C (grey), O (red), N (blue) and S (orange).

4.6 Discussion

The conjugation of SUMO onto target proteins is a reversible post-translational modification that regulates many processes including gene expression, cell cycle and stress responses (Cheng *et al*., 2006). SENP1 is a cysteine protease that catalyzes two essential reactions in the SUMO pathway (Yeh, 2009). It processes SUMO precursors to their mature form as well as deconjugates SUMO from the target proteins. Over-expression of SENP1 mRNA and protein level is observed in over half of the prostatectomy cases with high-grade prostatic intraepithelial neoplasia and in prostate cancer cells (Cheng *et al*., 2006; Li *et al*., 2012). The effect of its up-regulation on progression of prostate cancer has been confirmed in the transgenic mice (Chen *et al*., 2012). Therefore, SENP1 could be a promising therapeutic target for prostate cancer.

In this study, the fragment screen was performed with SENP1 using DSF and ligand-observed 1D NMR techniques. A commercial library of approximately 1000 fragments was screened with DSF at 1 mM concentration, but the screen did not identify any hits. The screening concentration was subsequently increased to 10 mM, and this resulted in 27 fragments exhibiting thermal shifts above a set threshold temperature of 2 ^{o}C. However, the protein stabilizing effects at such high compound concentrations were still considered to be insignificant and therefore their characterisation was not pursued further. Together with results observed for Siah1 (see section 3.3.1), these DSF data show that the technique is not amenable for screening fragments against such protein targets. DSF is not sensitive enough to reliably detect the binding of compounds with K_d above 1 mM (Hubbard and Murray, 2011).

SENP1 was also screened with 384 compounds from an in-house library of 1100 fragments (AstraZeneca) using ligand-observed NMR method. The library was specifically tailored for NMR screening and contained fragments of lower MW

compared to an SPR library used in the Siah1 screen. The NMR method is able to detect compound binding with affinities as low as 10 mM and has been successfully utilised for fragment screening of many protein targets (Brough *et al*., 2009; Chen and Hubbard, 2009). It is a relatively quick method, where fragments can be screened simultaneously in cocktails of 6-12 to achieve a higher throughput. In addition, information about the binding mode and fragment affinity can be obtained in cases where the site-specific inhibitor of the target protein is available. However, no reversible inhibitors of SENP1 are currently available commercially and no binding was observed with the four-residue peptide containing the C-terminal sequence of SUMO1. Thus, the NMR screen was performed without a positive control and the precise K_d values for the top fragments could not be measured, but were estimated to be in the millimolar range. The failure to identify any hits with DSF and NMR screens support the earlier notion that SENP1 is a challenging target for high-throughput and fragment screening. Future experiments might concentrate on a different set of fragments and other biophysical methods like SPR, which identified fragment hits for Siah1. For example, the fragment library could be modified to include chemical moieties resembling amino acids as many SENP1 inhibitors identified using *in silico* screening methods contain peptidomimetic scaffolds (Madu *et al*., 2013; Chen *et al*., 2012). Any future derived hits could be validated with fluorescence resonance energy transfer activity assay, which is well established for catalytic characterisation of SENPs (Shen *et al*., 2006).

In addition, the apo-structure of SENP1 was solved to 2.4 Å. The structure had a similar fold to previously characterised SENP1 proteins. However, a new feature was identified in the asymmetric unit. A Co^{2+} ion was assigned, which contributed to intermolecular contacts forming the crystal lattice.

Chapter 5

Results and discussion: Siah1 interaction studies with its binding proteins

5.1 Background

Siah1 ubiquitinating activity regulates over 30 proteins (**Table 1.1**) by targeting them for proteasomal degradation (see section 1.1.4). Siah1 acts either alone or as part of a multi-protein complex that is similar to APC and Skp1/Cul1/F-box protein ligases (Santelli *et al*., 2005). Complex assembly is thought to help recruit Siah1 substrates. In addition, a number of proteins interact with Siah1 to either block its activity or to stabilise Siah1 by inhibiting its auto-ubiquitination and degradation (Qi *et al*., 2013).

5.1.1 Siah1 substrates

Four Siah1 substrates were investigated. They were CtBP-interacting protein (CtIP), tribbles homolog 3 (TRB3), kinesin like DNA binding protein (Kid), and prolyl hydroxylase 3 (PHD3). These particular proteins were chosen based on the available information supporting their interaction with Siah1 and the results of bioinformatic analysis. CtIP is a transcriptional co-repressor that regulates cell cycle progression and associates with BRCA1 to mediate DNA double stranded break repair (Wang *et al*., 2012). CtIP was shown to interact with Siah1 both *in vitro* and *in vivo* (Germani *et al*., 2003). TRB3 is a presudokinase, acting as a scaffolding protein in insulin signalling and bone morphogenetic protein signalling pathways (Yokoyama and Nakamura, 2011). Its binding to Siah1 was identified using the yeast two-hybrid system and validated by co-immunoprecipitation assays (Zhou *et al*., 2008). Kid is implicated in the normal progression of mitosis and was observed to bind Siah1 in a yeast two-hybrid screen (Germani *et al*., 2000). PHD3 is involved in a cellular response to hypoxia, where it can catalyse the hydroxylation of HIF1α, leading to its degradation. PHD3 binding to Siah proteins was first identified by mass spectrometry analysis, and was shown to result in PHD3 turnover under hypoxic conditions (Nakayama *et al*., 2004). These proteins were suggested to interact with the SBD of Siah1 involving a conserved Val-x-Pro (VxP)

motif in their sequences, as was shown to be the case for Siah-interacting protein (SIP; House *et al*., 2006). However, with some proteins containing more than one VxP sequence the exact interaction sites and mode of action are not known.

5.1.2 Siah1 binding protein: SIP

SIP, also known as calcyclin binding protein, is a 228 residue (about 30 kDa) protein normally found in the cytoplasm, but which translocates to the nucleus upon increase in intracellular Ca^{2+} levels (Wu *et al*., 2003). SIP is an adaptor protein that mediates its function through protein-protein interactions; with roles in ubiquitination, proliferation, cytoskeletal rearrangement and regulation of transcription (Schneider and Filipek, 2011). SIP was identified to bind Siah1 as part of a multi-protein complex that also includes Skp1 and Ebi to regulate β-catenin levels (Matsuzawa and Reed, 2001). The interaction was subsequently confirmed by a number of biophysical studies (Santelli *et al*., 2005; Bhattacharya *et al*., 2005). Sequence analysis revealed SIP contains three distinct regions (Bhattacharya *et al*., 2005). The N-terminal part forms a helical hairpin domain involved in homodimerisation, followed by an unstructured sequence involved in binding Siah1. The central part folds into a seven stranded mixed β-sheet, important for binding Skp1. Finally, the disordered C-terminal segment is involved in binding calcyclin (Schneider and Filipek, 2011). The ligand-observed NMR studies using a shorter 80 residue isoform of SIP (SIP-S), containing identical 72 amino acids at the N-terminus to the full length SIP, mapped the Siah1 interaction sequence to ^{60}PAAVVAP^{66} (Santelli *et al*., 2005). The K_d of SIP-S for Siah1 was measured by ITC and NMR as 10 ± 5 μM. The same group also determined the structure of Siah1 in complex with a 13 residue peptide of SIP-S, thereby revealing the binding site on Siah1 (**Figure 1.5**). In this structure, the ^{60}PAAVVAP^{66} motif forms a β-strand that lies parallel to the β-sheet of SBD of Siah1. The conserved valine in a VxP motif packs

against Phe165 and Val179, while the conserved proline stacks against the side chain of Trp178. Although this study confirmed SIP-Siah1 complex formation, it is not clear whether other parts outside the N-terminal region of SIP also contribute to the binding. Thus, further studies are needed using longer SIP constructs.

5.1.3 Siah1 binding protein: PEG3

Paternally expressed gene 3 protein (PEG3) is a large multi-functional protein of nearly 1600 amino acids (about 165 kDa). Its sequence contains twelve Cys_2-His_2 type zinc finger motifs, extended segments of predicted disorder and the N-terminal SCAN domain, named after the first letters of four founding members of the family (SRE-ZBP, Ctfin51, AW-1 (ZNF174), and Number 18) (Williams *et al*., 1995). PEG3 has predominantly nuclear localisation and is able to bind DNA in a sequence specific manner, suggesting it can act as a transcription factor (Relaix *et al*., 1996; Thiaville *et al*., 2013). In addition, PEG3 is known to interact with proteins in the cytoplasm to regulate tumour necrosis factor and Wnt (the name is a combination of homologous genes Wg (wingless) and Int) signal transduction pathways (Relaix *et al*., 1998; Jiang *et al*., 2010). The initial binding of PEG3 to a Siah2 construct missing the RING domain was detected in a yeast two-hybrid screen (Relaix *et al*., 2000). The same study reported the interaction of a mouse PEG3 with both Siah proteins by immunoprecipitation, while the later experiments using deletion generated constructs of human PEG3 revealed that the N-terminus including the SCAN domain (residues 1-268) were required for binding the full-length Siah1 (Jiang *et al*., 2010). Interestingly, the SCAN domain of PEG3 (PEG3-SCAN) contains the amino acid sequence of ^{58}VGP^{60}, which could be a consensus Siah1 binding motif, but studies are required to confirm this hypothesis.

5.2 Aims

The aims of this project were to validate the binding of the aforementioned proteins to Siah1, and to elucidate the interaction sites and mechanisms. Initial objectives were to express and purify soluble proteins for subsequent interaction studies using size exclusion chromatography (SEC). In addition, the crystal structures of apo-proteins and in complex with Siah1 were sought to characterise interaction interfaces and gain a better understanding of their cellular roles.

5.3 Interaction studies

5.3.1 Substrates

Protein constructs of CtIP, TRB3, Kid, and PHD3 (**Table 5.1**) were designed according to the published information on Siah1 interaction regions and also bioinformatic analysis using web-based services for protein structure prediction Phyre2 and I-TASSER (Kelley and Sternberg, 2009; Zhang, 2008). The genes encoding these constructs were purchased from a commercial vendor and were then inserted into a pET15b-TEV vector for protein expression following the same protocol as for Siah1 (see section 2.2-2.4). CtIP and TRB3 expressed insolubly (**Table 5.1**), even when expressed together with the N-terminally attached solubility enhancing proteins such as maltose-binding protein and glutathione-S-transferase. Kid and PHD3 had a low yield, exhibited precipitation and interacted with *E.coli* proteins, resulting in insufficient final yield for further experiments with Siah1. Therefore, the binding of these proteins to Siah1 could not be investigated further.

Protein	UniProt No.	Constructs	Expression
CtIP	Q99708	668-897	Insoluble
TRB3	Q96RU7	66-358 / 66-318	Insoluble
Kid	Q14807	435-665	Low/ impure
PHD3	Q9H6Z9	1-239	Low/ precipitates

Table 5.1 Protein constructs and their expression.

5.3.2 SIP

Full-length SIP was expressed and purified as described in sections 2.3 and 2.4. SIP eluted from a size exclusion column as a single species displaying a mass of about 94 kDa (**Figure 5.1**). This value equates to 3.6 times the MW of the monomer, which is consistent with previous gel filtration studies that reported SIP self-association (Matsuzawa and Reed, 2001). The crystal structure of SIP-S subsequently revealed that SIP forms a dimer (Santelli *et al*., 2005). SIP possibly displays a higher MW on a gel filtration column due to its elongated rod like shape. However, one study also observed the presence of monomer under certain buffer conditions (Bhattacharya *et al*., 2005). The purified SIP was investigated with DSF, which suggested the protein was folded and displayed the highest melting temperature of 61 ºC in the buffers tested. SIP interaction with Siah1 was tested by combining proteins together at 1:1 stoichiometry and incubating the sample overnight prior to its analysis on a size exclusion column. The results are shown in **Figure 5.1**. GF chromatogram showed the presence of two peaks. SDS-PAGE analysis revealed that the first peak of higher MW contained both proteins. The second peak with approximate mass of 49 kDa contained solely Siah1. The presence of Siah1 in fractions eluting at higher MW indicates the formation of complex with SIP. However, the ratio of each molecule in a complex could not be determined. For example, the Siah1 dimer binding two molecules of SIP would result in a complex with MW of 97 kDa, whereas the peak conformed to assembly with

approximate mass of 133 kDa. In addition, the bands on SDS-PAGE gel corresponding to SIP and Siah1 are not proportional, suggesting the sample is not homogeneous. Siah1 interaction with SIP might be transient, where equilibrium is established between Siah1-SIP complex and SIP homodimer. Attempts to crystallise SIP individually and in complex with Siah1 proved unsuccessful.

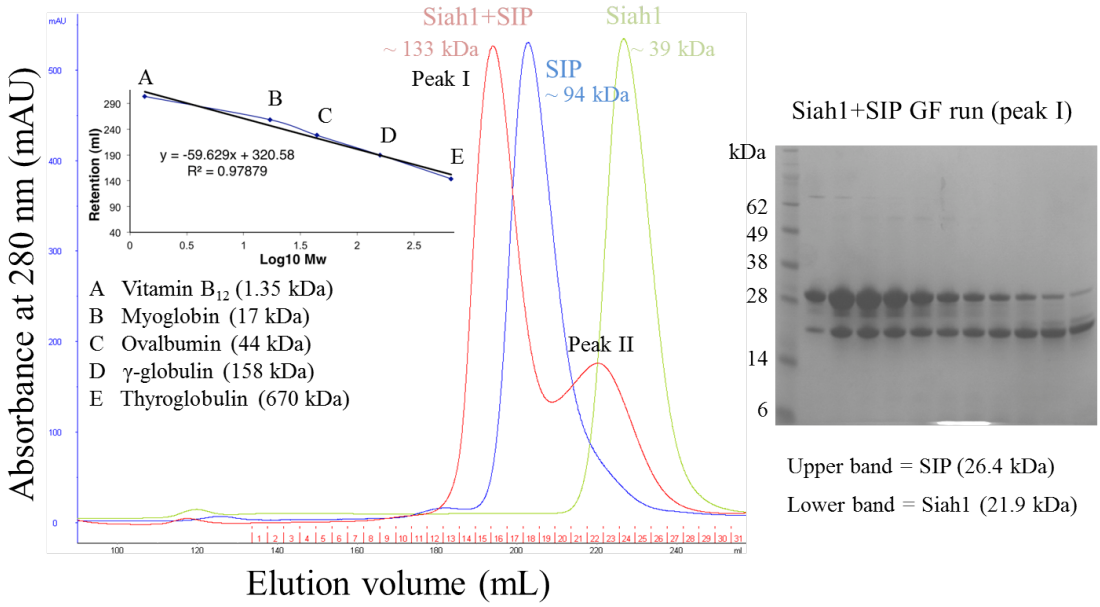

Figure 5.1 Size exclusion chromatography for Siah1 and SIP.
Three GF runs are superimposed: Siah1+SIP (red), SIP (blue), Siah1 (Green). SDS-PAGE gel is showing fractions from peak I of Siah1 run together with SIP. Siah1 elution volume is decreased in the presence of SIP compared to when run individually, thereby indicating formation of Siah1-SIP complex.

5.3.3 PEG3-SCAN

The gene encoding the SCAN domain of PEG3 (amino acids 40-130) was sub-cloned into the pET15b-TEV plasmid and transformed into *E.coli* for protein expression (see sections 2.2 and 2.3). The protein was purified according to the protocol in section 2.4. The protein eluted from a GF column as a single species with a mass of approximately 30 kDa. The theoretical mass of PEG-SCAN is 11.4 kDa; thereby its oligomeric state could not be determined from a GF run. DSF analysis suggested the presence of a globular SCAN domain, which displayed a melting temperature of 52 ºC in a number of buffers. Since no buffer appeared to enhance stability the protein was left in the GF

buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl). The association between PEG3-SCAN and Siah1 was tested in SEC by combining protein samples together at 1:1 stoichiometry. The mixture was left overnight at 4 ^{o}C, before it was run on a GF column. However, there was no evidence of complex formation (**Figure 5.2**). In addition, the interaction was tested using protein-observed NMR, which was performed by Dr. Navratna Vajpai (AstraZeneca). The NMR experiment was done under the following conditions, where 100 µM of ^{15}N-labeled Siah1 was mixed with 100 µM of unlabeled PEG3 in 50 mM HEPES pH 7.5, 50 mM NaCl, and 5 % D_2O buffer. The chemical shift perturbations in the 2D ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled Siah1 were monitored upon addition of PEG3. The data showed no differences in chemical shifts in the ^{1}H-^{15}N HSQC spectra once PEG3-SCAN was added (data not shown), thereby indicating an absence of interaction between the proteins. The PEG3-SCAN sample was used for subsequent crystallisation, as its structure might give insight into the role this domain plays within the PEG3 protein.

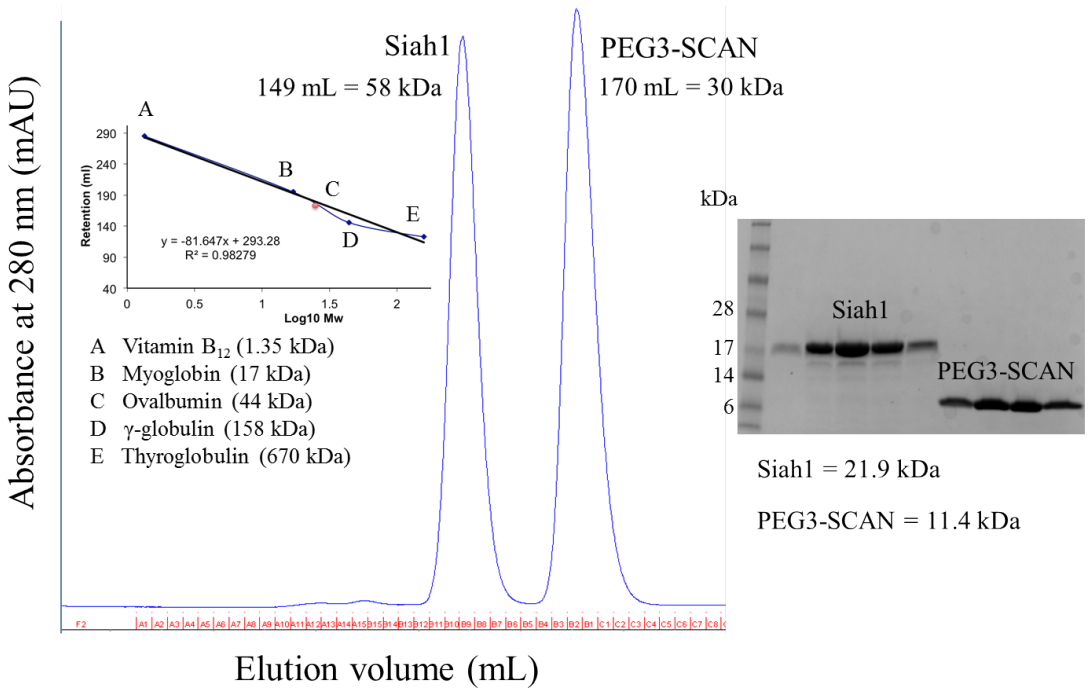

Figure 5.2 Size exclusion chromatography for Siah1 and PEG3-SCAN.
Sample containing Siah1 together with PEG3-SCAN at 1:1 stoichiometry was run on size exclusion column. Siah1 and PEG3-SCAN eluted in two separate peaks. Collected fractions were analysed on SDS-PAGE gel.

5.4 PEG3-SCAN crystallisation and data collection

Crystallisation screening of PEG3-SCAN was carried out with the high-throughput Phoenix liquid handling system (Art Robbins Instruments/ Rigaku) and several commercially available screens (see section 2.5.3). 100 nL of protein sample at ~16.5 mg mL^{-1} were mixed in sitting-well plates with 100 nL of reservoir solution against 70 μL of the same reservoir solution. Crystals grew within a day in many conditions that contained polyethylene glycol (PEG) of different molecular weight. These crystallisation conditions were scaled up from these nano-drops to micro-drops of total volume 4 μL. However, the rod-shaped crystals that were observed were small, approximately 0.1 x 0.02 x 0.02 mm, and gave poor diffraction. While preparing more protein for use in crystal optimization it was observed that larger crystals actually formed spontaneously when the protein was concentrated in the GF buffer using Vivaspin 20 concentrators with a 3,000 MW cut off (Sartorius Stedim Biotech). Hexagonal bipyramid crystals, reaching 0.2 x 0.2 x 0.2 mm dimensions, formed within minutes (**Figure 5.3**). The average protein concentration in the centrifugal device was 5 mg mL^{-1}, but likely to have been considerably higher near the membrane where crystal nucleation occurred. The selection of a suitable cryoprotectant required extensive screening and optimization. The use of glycerol, ethylene glycol and paratone-N produced either poor diffraction or pronounced ice rings. The most favourable cryoprotectant was PEG200. Crystals were transferred into cryo-solution of PEG200 and GF buffer at 1:1 ratio prior to flash cooling for X-ray diffraction studies.

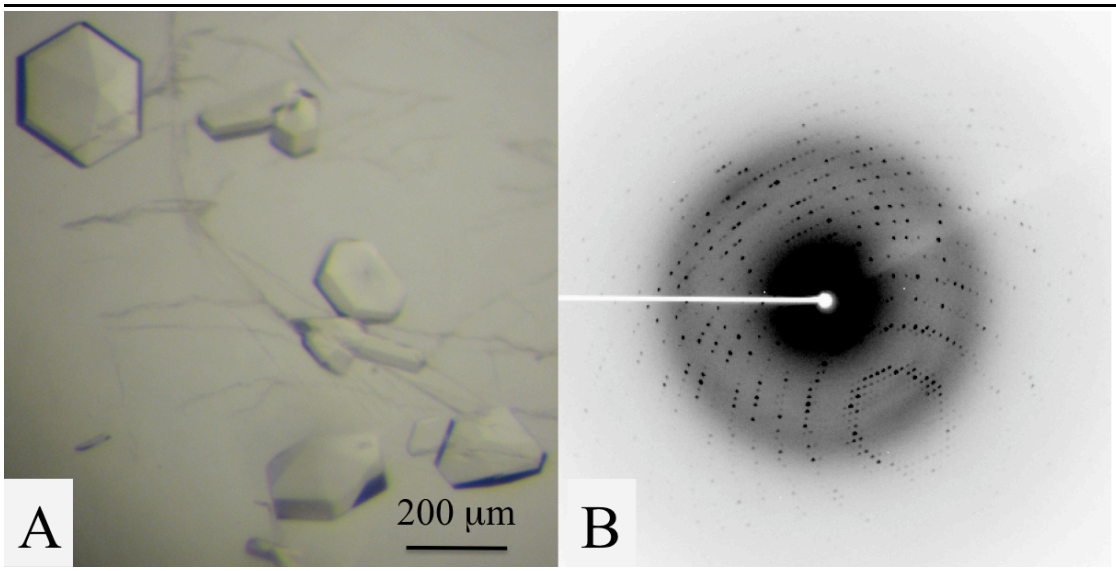

Figure 5.3 PEG3 crystals.
(A) The crystals formed during protein concentration within minutes. (B) The crystals diffracted in-house to 1.95 Å.

5.4.1 Structure solution and refinement

Diffraction data were collected in-house with a Micromax-007 rotating anode generator (copper K_α, λ = 1.5414 Å) coupled to an AFC11 Saturn 944+ CCD detector (Rigaku). The data were indexed and integrated with *iMOSFLM* (Battye *et al*., 2011) and scaled with *AIMLESS* from the *CCP4* program suite (Winn *et al*., 2011). The structure was solved by molecular replacement with *Phaser* (McCoy *et al*., 2007) using a poly-Ala model of the SCAN domain dimer from the mouse zinc finger protein 206 (Zfp206, PDB code 4E6S (Liang *et al*., 2012)) that shares 38 % sequence identity with the PEG3-SCAN domain. The output model was subjected to a round of rigid body and restrained refinement using *REFMAC5* (Murshudov *et al*., 2011). The poly-Ala model was modified to the sequence of human PEG3-SCAN based on inspection of electron and difference density maps in *COOT* (Emsley *et al*., 2010). Several residues and side chains for which there was no convincing electron density were deleted. Additional rounds of restrained least-squares refinement followed, interspersed with map inspection and model manipulation. The refinement used the default geometry and *B*-

factor restraint weights. Neither non-crystallographic symmetry (NCS) restraints nor TLS (Translation/ Libration/ Screw) were used in refinement. A number of ligands (ethylene glycol, diethylene glycol and triethylene glycol) were included in the model on the basis of the difference density and chemical environment, and refined successfully. These molecules are likely to be decomposition products or impurities of the PEG200 cryoprotectant. The final model also includes multiple side chain conformers and water molecules. The stereochemistry of the model was checked using *MolProbity* (Chen *et al*., 2010). Crystallographic statistics are given in **Table 5.2**.

	PEG3-SCAN
Space group	$P6_5$
Unit cell dimensions: *a*, *b*, *c* (Å)	83.61, 83.61, 55.23
Resolution[a] (Å)	13.8 - 1.95 (2.00 - 1.95)
No. of reflections	453776 (23734)
Unique reflections	16090 (1143)
Completeness (%)	99.7 (99.8)
Multiplicity	28.2 (20.8)
$<I/\sigma I>$	38.2 (9.7)
Wilson *B* (Å^2)	20.6
Residues	
Chain A	40-127
Chain B	40-129
Water/ ethylene glycol/ diethylene glycol/ triethylene glycol	155 / 21 / 4 / 2
R_{merge}[b] (%)	7.0 (31.8)
R_{work}[c], R_{free}[d] (%)	17.2 / 22.4
Mean *B*-factors (Å^2)	
Protein	26.6
Ligands and waters	40.6
Ramachandran plot[e] (%)	
Most favoured	98.4
Additional allowed	1.6
Outliers	0.0
R.m.s.d. on ideal values[f]	
Bond length (Å) / bond angles (°)	0.02 / 2.02

Table 5.2 Crystallographic statistics of PEG3-SCAN.
a. Values in parentheses refer to the highest resolution shell. **b.** $R_{\text{merge}} = \sum_{hkl}\sum_i |I_i(hkl) - <I(hkl)>| / \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $<I(hkl)>$ is the mean value of $I_i(hkl)$ for all *i* measurements. **c.** $R_{work} = \sum_{hkl}||F_o|-|F_c|| / \sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. **d.** R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from refinement calculations. **e.** (Chen *et al*., 2010). **f.** (Engh and Huber, 1991).

5.5 PEG3-SCAN structure

5.5.1 Structure quality

Crystals of human PEG3-SCAN belong to space group $P6_5$, with a V_M value of 2.44 $Å^3$ Da^{-1} and solvent content of approximately 50 % for an asymmetric unit comprising two polypeptide chains. The polypeptides are arranged as a symmetrical dimer consistent with previously solved structures of SCAN domains (Liang *et al*., 2012; Ivanov *et al*., 2005; Peterson *et al*., 2006). The crystals diffract to a resolution of 1.95 Å and continuous, well-defined electron density is observed for majority of the structure, with an exception of a few residues at the C-terminus. The final model also contains two extra residues (His and Met) at the N-terminus, which were left from a cleaved histidine tag. A Ramachandran plot indicates that 98.4 % of the amino acids are located in the most favoured region with no outliers.

5.5.2 Overall structure

The human PEG3-SCAN domain resembles a broadened V-shaped structure (**Figure 5.4A**), with approximate dimensions of 50 x 25 x 25 Å. Two V-shaped monomers in a dimer are located side by side to form a dimer with dimensions of 50 x 37 x 30 Å. The subunit consists of five α helices, with primary sequence and assigned secondary structure displayed in **Figure 5.4B**. The N-terminal sub-domain is formed from helices α1 and α2 which form one of the two arms of the V (**Figure 5.4A**). The other half, the C-terminal domain packs together α3, α4 and α5. The domain assemblies with the N-terminal sub-domains interacting with partner C-terminal sub-domains (**Figure 5.5**). A least-squares superposition of subunits with *LSQKAB* (Kabsch, 1976) gives the r.m.s.d. (root-mean-square deviation) of 0.57 Å for 90 Cα atoms, which shows there are no major conformational differences between the two subunits.

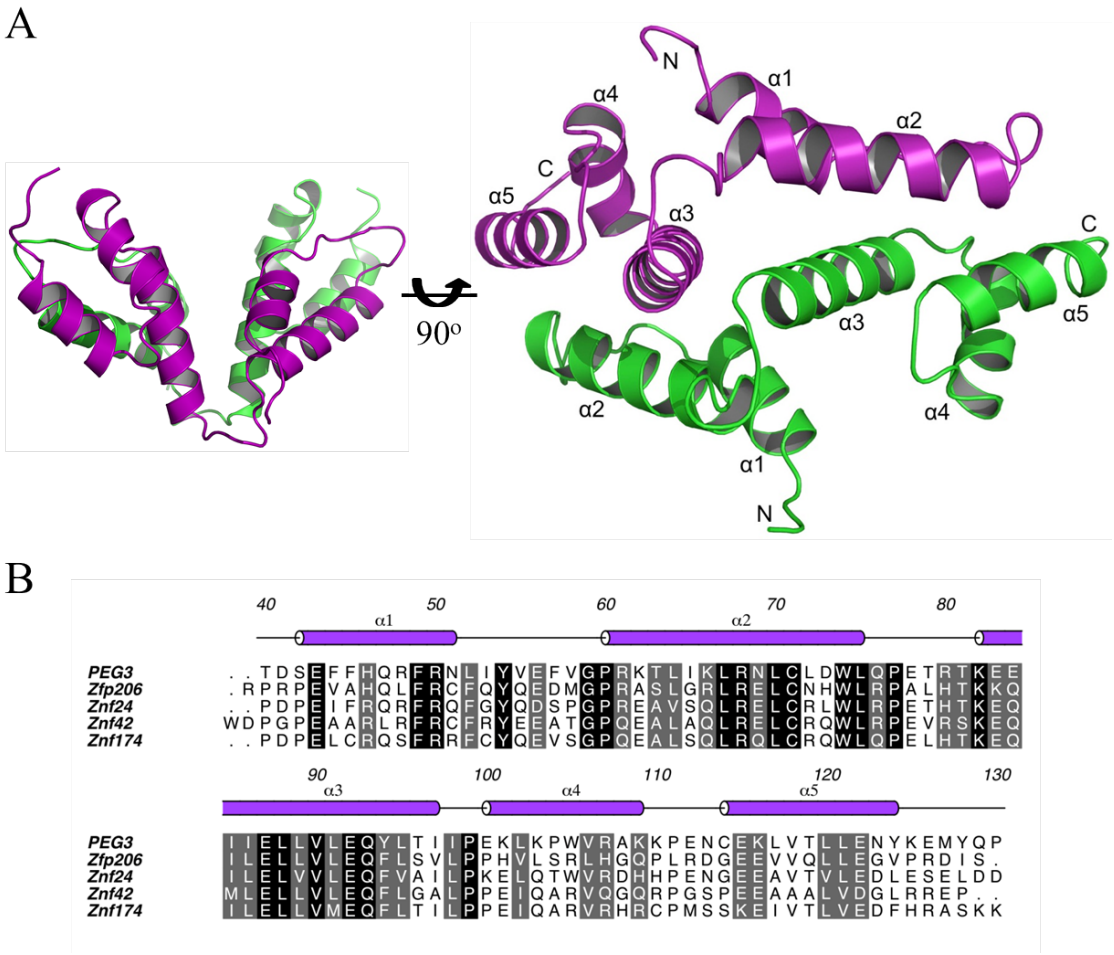

Figure 5.4 The structure of PEG3-SCAN.
(A) Overall structure of PEG3-SCAN homodimer is shown as ribbons with one subunit green, the partner purple. The N- and C- termini as well as the five α-helices of each monomer are labelled accordingly. (B) Diagram is showing sequence alignment of PEG3-SCAN with other SCAN proteins from PDB. PEG3-SCAN residues that are strictly conserved in Zfp206 (PDB: 4E6S), Znf24 (PDB: 3LHR), Znf42 (PDB: 2FI2) and Znf174 (PDB: 1Y7Q) are encased in black, while residues sharing similar properties in five proteins are encased in grey. Five α-helices are shown as cylinders (purple) and are numbered accordingly. The numbers that are shown above the secondary structure mark residues in the full length PEG3 protein.

The total accessible surface area of the individual subunits as determined by *PISA* (Krissinel and Henrick, 2007), is about 7400 Å^2 of which 1700 Å^2 involved in forming dimer interface. Thus, approximately 23 % of the surface area of each subunit participates in dimerisation. *PISA* estimates the free energy of dissociation (ΔG^{diss}) to be 19.4 kcal mol^{-1}, suggesting that the assembly is thermodynamically stable. This is consistent with the observation of a stable oligomer in solution. These surface areas are equivalent in other SCAN structures. For instance, the interface area and ΔG^{diss} for the Znf24 dimer (PDB code 3lhr) are 23 % and 21.8 kcal mol^{-1}, respectively.

Figure 5.5 Overlay of SCAN domain structures.
These are PEG3-SCAN (purple), Zfp206 (PDB 4e6s, cyan), Znf24 (PDB 3lhr, green), Znf42 (PDB 2fi2, yellow) and Znf174 (PDB 1y7q, grey).

There are currently four SCAN domain structures deposited in the PDB, two crystal structures and two determined by solution NMR. Their sequence conservation with PEG3-SCAN is presented in **Figure 5.4B**. A large degree of structural conservation is observed between these SCAN domain structures and PEG3-SCAN (**Figure 5.5**), with calculated r.m.s.d. values shown in **Table 5.3**. Two regions of structural non-alignment are at the N- and C-terminal ends, which are more flexible than the core of the molecule. The other flexible segment is α4, which is positioned away from the dimer interface and whose sequence is least conserved among five α-helices. The r.m.s.d. values for overlaid PEG3-SCAN with the two structures derived by NMR (Znf42 and Znf174) are higher than for the X-ray structures. This is because the NMR structures contain greater uncertainties and 20 conformers were averaged during superimposition with PEG3-SCAN domain.

Protein name	PDB codes	R.m.s.d (Å)	R.m.s.d alignment length	Sequence identity (%)
Zfp206	4E6S	1.57	157	38
Znf24	3LHR	1.51	164	48
Znf42	2FI2	2.85	155	35
Znf174	1Y7Q	2.87	167	43

Table 5.3 Structure and sequence similarity of PEG3-SCAN and other SCAN domains.

These included crystal structures of Zfp206 and Znf24, and solution NMR structures of Znf42 and Znf174. R.m.s.d. calculations were carried out with PDBeFold using secondary structure matching (Krissinel and Henrick, 2004) with the PEG3-SCAN dimer in the superposition. Sequence alignment was performed with ClustalW2 using residues 40-130 of the full-length PEG3 against the core of the SCAN domain, as well as 2-5 flanking residues, of other proteins.

5.5.3 Residues forming the SCAN dimer interface

The PEG3-SCAN homodimer consists of an extended network of interactions, such as hydrogen-bonding, salt-bridge interactions and van der Waals forces, which weave together both subunits. The residues found at the dimer interface in all five structures are highly conserved, even though the overall sequence conservation among SCAN domains is only 40-50 % (**Figure 5.4B**). The intermolecular interactions are predominantly formed by residues located on α1 and α2 of the N-terminal segment of one subunit and amino acids on α3 of the C-terminal subdomain of the second. Sequence homology is highest between helices α2 and α3, with their residues participating in either direct inter-subunit contacts or mediating the chains of hydrogen bonding that involve a number of residues. For instance, the carboxyl group of Glu43 donates a proton to stabilise the imidazole ring of His46 located on the same helix α1 (**Figure 5.6A**). The later makes another hydrogen bond to Glu92 on α3, which is located close to the NCS two-fold axis. Glu92 in turn forms a salt bridge with Arg50 and another bond with well-ordered and buried water molecule. The Arg50 also interacts with Gln93 of the partner subunit and a water molecule, which acts as a bridge to link Arg50 to the partner subunit Glu92. Arg68 forms an inter-molecular salt bridge with Glu87, linking α2 and α3 from two different subunits together (**Figure 5.6B**). The aforementioned residues are highly conserved within SCAN domains and observed to

establish similar hydrogen bonding contacts. In addition, mutation of both the equivalent Arg50 and Arg61 residues to alanines in Zfp206 destabilizes its heterodimerisation with Zfp110 (Liang *et al*., 2012). Thus, these invariant residues seem to play an important role in both homo- and heterodimerisation. Other residues involved in inter-molecular bonding include Lys82, whose side chain interacts with Pro77 and Arg61 with Glu115 (**Figure 5.6B**). The later makes a salt-bridge interaction, which links α2 with the partner subunit α5. The later pair of residues is substituted to a similar lysine glutamine pair in some SCAN domain sequences.

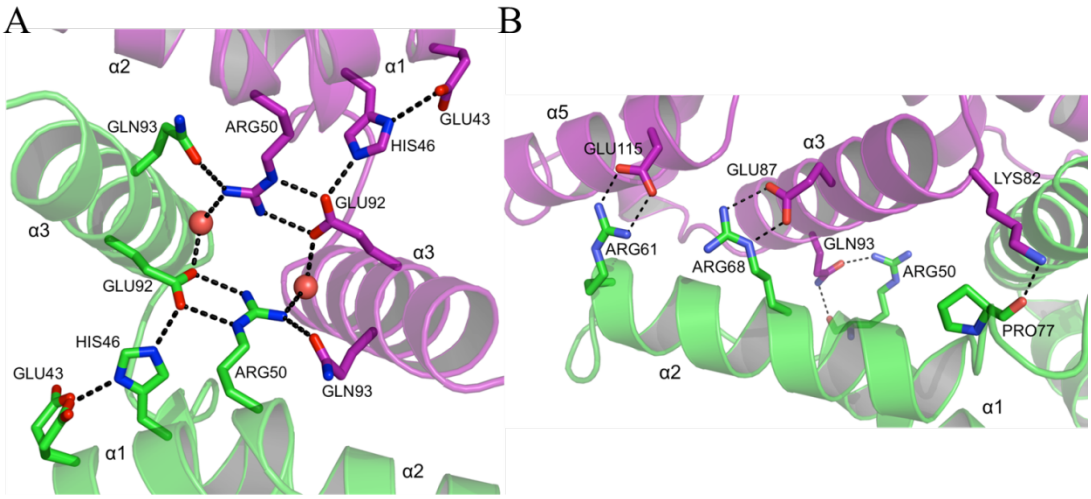

Figure 5.6 The dimer interface of PEG3-SCAN (I).
(A) A network of hydrogen-bonds is established between conserved residues lining the dimer interface. (B) A second cluster of hydrogen bonding and salt bridge interactions at the dimer interface. Water molecules are shown as red spheres. Atomic positions are coloured: C (purple or green), O (red), and N (blue).

A number of the interactions present in the PEG3-SCAN dimer are absent in other structures. For example, a hydrogen bond is formed between the side chain of Tyr94 on one subunit and the main chain of Pro60 (**Figure 5.7A**) on the partner. This interaction does not happen in other structures where the tyrosine is swapped for phenylalanine. In addition, PEG3-SCAN establishes a salt bridge between Glu56 and non-conserved Lys101. Glutamate is present at this position in most other SCAN domains. Such subtle structural variations may confer the preference of different SCAN domains to form

homo- and heterodimers.

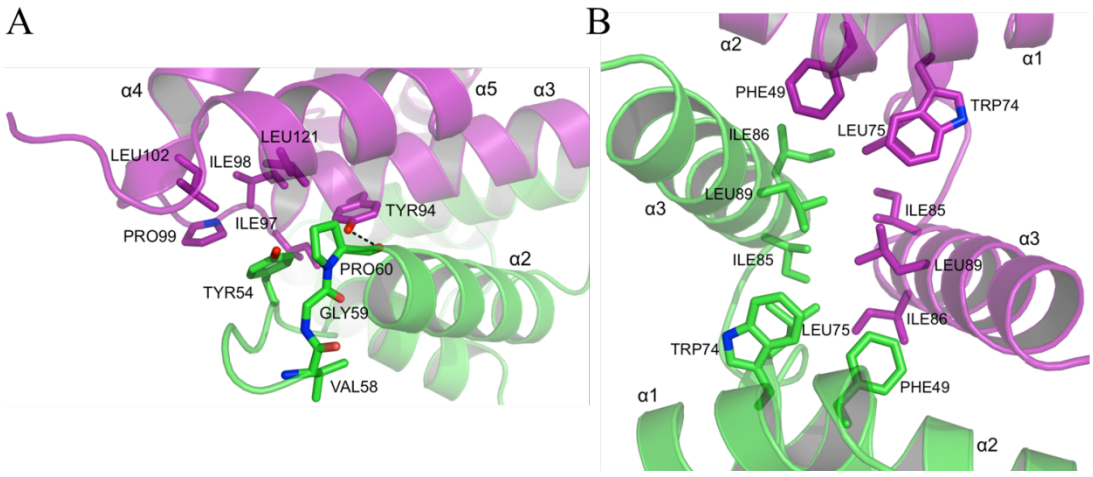

Figure 5.7 The dimer interface of PEG3-SCAN (II).
(A) A hydrophobic patch is present at each end of the assembly to stabilize the dimer. The conserved Tyr94 extends across the dimer interface, contributes to hydrophobic interations and makes a hydrogen bond with the carbonyl of Pro60. (B) A group of conserved, aliphatic and aromatic residues form a hydrophobic core at the centre of the dimer.

Further to the water molecules described above, with their central location to mediate a number of inter-subunit interactions, other water molecules are positioned at the periphery of the dimer interface. These seem to be less ordered and likely do not contribute significantly towards stabilisation of the dimer, taking into consideration the high percentage of the surface area involved in direct association.

The hydrogen interactions are further strengthened by hydrophobic contacts. For example, one hydrophobic patch includes Leu75, Ile85, Ile86 and Leu89, which are located on α2 and α3, and make interactions to equivalent residues in opposing monomer (**Figure 5.7B**). In another patch, a conserved Tyr54 in one monomer buries its side chain into a hydrophobic core of Tyr94, Ile97, Ile98, Pro99, Leu102 and Leu121 of α3-α5 of the second monomer in a dimer (**Figure 5.7A**). Other hydrophobic residues extending across the dimer interface include Leu52, Leu64, Leu67, Val90 and Val118. In addition, the superposition of SCAN domain structures also helps explain the role of other conserved residues. For instance, the strictly conserved Pro60 and Pro98 introduce

kinks at the start of α2 and α4, respectively. Thus, the sequence alignment and structural overlay of PEG3-SCAN with available SCAN domain structures shows overall fold homology and reveals the key residues for forming PEG3-SCAN dimer.

5.5.4 Function of the SCAN domain

Some transcription factors commonly form complexes to control the expression of target genes. For instance, the proto-oncogene Jun, which is a member of the basic region-leucine zipper (bZIP) family, controls gene transcription in a precise way by forming homo- and heterodimers. The dimerisation mode and interactions with given binding partners regulate its activity and DNA-binding site preferences (Halazonetis *et al*., 1988). Certain Cys_2-His_2 zinc finger containing transcription factors are also known to be regulated by a similar mechanism via their protein-protein binding BTB/POZ domain. Self-oligomerisation of this domain in the GAGA transcription factor promotes DNA binding affinity and specificity (Espinas *et al*., 1999). The SCAN domain may have a similar function in regulating mammalian gene expression by facilitating both homo- and specific hetero-dimerisation of Cys_2-His_2 zinc finger proteins. Currently there is no evidence that would indicate SCAN domain's involvement in direct transcriptional activation or repression. The ability of PEG3-SCAN to homodimerise *in vitro* provides the possibility that the same can happen with full-length protein inside the cell. Therefore, it might bind to at least some target genes as a homodimer. In particular, formation of the homodimer might enable PEG3 to bind palindromic DNA motifs. The support for this hypothesis comes from the transcription factor Zfp206 that can form a homodimer using its SCAN domain for binding to a palindromic sequence (Liang *et al*., 2012; Yu *et al*., 2009). A search for PEG3 DNA interaction motifs revealed a 15 base pair non-palindromic sequence, which was subsequently validated by electromobility shift and promoter assays (Thiaville *et al*., 2013). Even though only the

full-length motif exhibited the maximal binding, PEG3 was also detected to interact with the partial sequence as well as to other segments with degenerate sequence.

There are currently four known isoforms of PEG3. It is possible that isoform specificity affects the DNA binding selectivity and affinity. For example, one isoform contains only two Cys_2-His_2 zinc finger motifs, while in two others the SCAN domain is absent altogether. Thus, cellular expression of PEG3 with a different number of zinc finger motifs as well as with and without the SCAN domain can potentially act as another level of regulation. A number of studies have shown that at least some SCAN domain members are able to selectively bind to other SCAN members (Williams *et al*., 1999; Sander *et al*., 2000; Schumacher *et al*., 2000; Liang *et al*., 2012). However, the possibility of PEG3-SCAN domain to heterodimerise has not been investigated to date. Residues participating in inter-subunit binding in the Zfp206 dimer structure were closely matched and had similar spatial location to the residues of an overlaid PEG3-SCAN subunit. This indicates that PEG3-SCAN domain might form heterodimers with Zfp206 and with other SCAN domain containing proteins. Future studies, investigating the binding partners of PEG3-SCAN are needed to help determine the function of PEG3 in gene regulation.

5.6 Discussion

Siah1 is a scaffolding protein reported to interact with a large number of different proteins (Qi *et al*., 2013). Many are the substrates for ubiquitination and subsequent proteasomal degradation. Others regulate Siah1 activity and some assemble together with Siah1 into multi-protein E3 ligase complexes. In attempt to validate these and characterise the relevant protein-protein interactions, a number of Siah1 substrates and its binding partners in multi-protein complexes were investigated. Initial studies concentrated on four target proteins: CtIP, TRB3, Kid and PHD3. However, attempts to

produce soluble proteins in *E.coli* in sufficient amounts for the interaction studies were unsuccessful. Future studies may try to co-express these proteins together with Siah1, explore a different expression system or use alternative protein constructs.

SIP was shown to interact with Siah1 in an E3 ligase complex, where it physically links the ubiquitination activity of Siah1 with Skp1 (Schneider and Filipek, 2011). The binding site for Siah1 is situated at the N-terminal segment of SIP, which also contains a Siah1 binding motif within its sequence (Santelli *et al*., 2005). This study aimed to assess whether other SIP segments outside the N-terminal domain were involved in the interaction with Siah1. The interaction between Siah1 and a full-length SIP was initially confirmed by the presence of a complex on a gel filtration column. However, subsequent attempts to crystallise SIP in complex with Siah1 and alone were unsuccessful. Crystallisation might be hindered by the transient nature of the SIP-Siah1 interaction or by the unstructured C-terminal region of SIP. The use of other components such as Skp1 and Ebi to stabilise an E3 ligase complex may be considered in future studies.

Two previous studies reported the interaction between PEG3 and Siah1 (Relaix *et al*., 1998; Jiang *et al*., 2010), and showed that they cooperate together to induce p53 independent apoptosis. However, the studies were not conclusive so the interaction was investigated here by SEC and NMR. However, no association was observed between the proteins either on a size exclusion column or in the 2D NMR experiment using ^{15}N-labelled Siah1. The absence of the interaction indicates that residues outside the SCAN domain might be needed for binding to Siah1. However, there are no other obvious Siah1 binding motifs nearby and the sequences on both ends of SCAN domain are predicted to be disordered. The PEG3-SCAN structure reveals that the ^{58}VGP^{60} motif is located on the loop just prior to α2 (**Figure 5.7A**). The proline is buried within a dimer

interface, while the side chain of valine is pointing to the side of α2. If this is indeed a Siah1 consensus site, then conformational changes might be required to allow for complex formation. It is possible the conditions under which the experiments were performed were not suitable for such changes to occur or the presence of the VxP sequence might be a coincidence. The data obtained cannot exclude the possibility that interaction occurs though other sites outside the SCAN domain, thereby additional studies are needed with longer fragments of PEG3.

The determined structure of the SCAN domain from PEG3 revealed it forms a stable homodimer. The structure shared high structural and sequence homology with other known SCAN domains and uncovered the key residues forming the dimer interface. The ability to dimerise suggests PEG3 might form homo- and heterodimers with other SCAN domain containing proteins to control gene expression in a combinatorial fashion. However, further studies are required to test this hypothesis.

Chapter 6

Results and discussion: *Burkholderia cenocepacia* cytosolic carboxypeptidase

6.1 Background

Metallocarboxypeptidases (MCPs) hydrolyse peptide bonds in target proteins to cleave C-terminal amino acids. They are classified into different families according to their sequence homology, catalytic mechanism and function (Gomis-Rüth, 2008; Rawlings *et al*., 2012). One of the more diverse families is called M14. Its members contain a single catalytic Zn^{2+} in the active site, which is held in place by two histidine residues, one glutamate and a water molecule. The family is grouped into four subfamilies known as M14A, B, C and D. One of the earliest and best studied family member being bovine carboxypeptidase A (CP-A; Christianson and Lipscomb, 1989). In general, M14A and M14B carboxypeptidases function either within the secretary pathway or are themselves secreted into the extracellular space. Members of the M14A subfamily are produced as inactive enzyme precursors that are activated following cleavage of N-terminal segments to reveal the active site. In contrast, the M14B subfamily members are produced as active enzymes with a transthyretin-like domain at the C-terminus. These proteins often contain additional domains and even repeats of the carboxypeptidase domain (Vendrell *et al*., 2004). The M14C group contains the bacterial orthologs of D-glutamyl-(L)-meso-diaminopimelate peptidase I and process components of the bacterial cell wall (Hourdou *et al*., 1993). They resemble the M14A group in that they also carry an N-terminal extension.

The M14D subfamily has only recently been described and there is still little information available about their structure and function. They are also known as cytosolic carboxypeptidases (CCPs) due to their cytosolic localisation. A search of available genomic sequences revealed CCPs are present in most organisms, with an exception of Archaea, Fungi and higher plants (Otero *et al*., 2012). The number of CCP paralogs varies between species. For example, six isoforms were identified in mammals,

two in nematodes and just one in proteobacteria, and only in Gram-negative bacteria containing flagella (Rodriguez de la Vega *et al*., 2007). Four mammalian CCP proteins and a homologue from *Caenorhabditis elegans*, called *Ce*CCPP-6, have been identified as deglutamylases (Kimura *et al.*, 2010; Rogowski *et al.*, 2010). They reverse glutamylation, process, a post-translational modification where glutamate residues are attached to a target protein. This modification was first observed on α- and β-tubulin, and later found to occur on telokin and myosin light chain kinase 1 (Edde *et al.*, 1990; Rudiger *et al.*, 1992; Rogowski *et al.*, 2010). Mammalian deglutamylases can be further distinguished by their substrate preference. For example, CCP1 and its functional homologs CCP4 and CCP6 cleave individual glutamates from the main chain, while the CCP5 isoform removes the branching point glutamates. In addition to this activity, CCPs might process other C-terminal amino acids. For example, mammalian CCP2 was recently reported to cleave the C-terminal tyrosine from α-tubulin (Sahab *et al*., 2010). Less is known about the functional activity of the bacterial CCPs. A recent study of the *Pseudomonas aeruginosa* CCP (*Pa*CCP) did not detect any activity against known carboxypeptidase substrates, with the authors suggesting that activation might require a conformational change or *Pa*CCP might be highly substrate specific (Otero *et al*., 2012). Although the function of bacterial CCPs is not understood, the gene encoding the *Pa*CCP was identified to be essential for the pathogen to establish a lung infection in a murine model (Winstanley *et al.,* 2009). Therefore, CCPs might represent a potentially novel class of antibacterial drug targets and require further investigation.

6.2 Aims

The aim of the study was to structurally characterise recombinant *Burkholderia cenocepacia* CCP (*Bc*CCP). *Bc*CCP is the closest ortholog of mammalian CCP5 and its function is unknown. The elucidation of its structure would allow us to compare *Bc*CCP

with *Pa*CCP (Otero *et al*., 2012) and two other bacterial CCP structures in the Protein Data Bank (PDB) together with mammalian carboxypeptidases of defined substrate specificity. Together they should provide an insight into the conserved residues and motifs forming a stable CCP fold and might inform about *Bc*CCP substrate specificity.

6.3 Recombinant *Bc*CCP expression and purification

The gene encoding the full-length protein (residues 1-384) was amplified using genomic DNA from *B. cenocepacia* J2315 (Laboratory for Microbiology of the Faculty of Sciences of the Ghent University (LMG) 16656 strain) and cloned into pET15b-TEV that incorporates an N-terminal histidine tag (see section 2.2). The protein was expressed in ArcticExpress (DE3) cells and was purified as described in section 2.4 still carrying the His-tag. The typical protein yields were in excess of 20 mg L^{-1} of culture. *Bc*CCP monomer has a theoretical MW of 45.3 kDa, but the protein eluted from a GF column with a mass of around 140 kDa (**Figure 6.1**). Native PAGE gel showed a band of around 146 kDa, thereby suggesting oligomer formation. The subsequent SEC-MALS experiment showed the protein eluted with a mass of 174 kDa (data not shown), indicating that *Bc*CCP forms a tetramer. Gene cloning, SEC-MALS and crystallisation experiments were carried out by Thomas Eadsforth (University of Dundee).

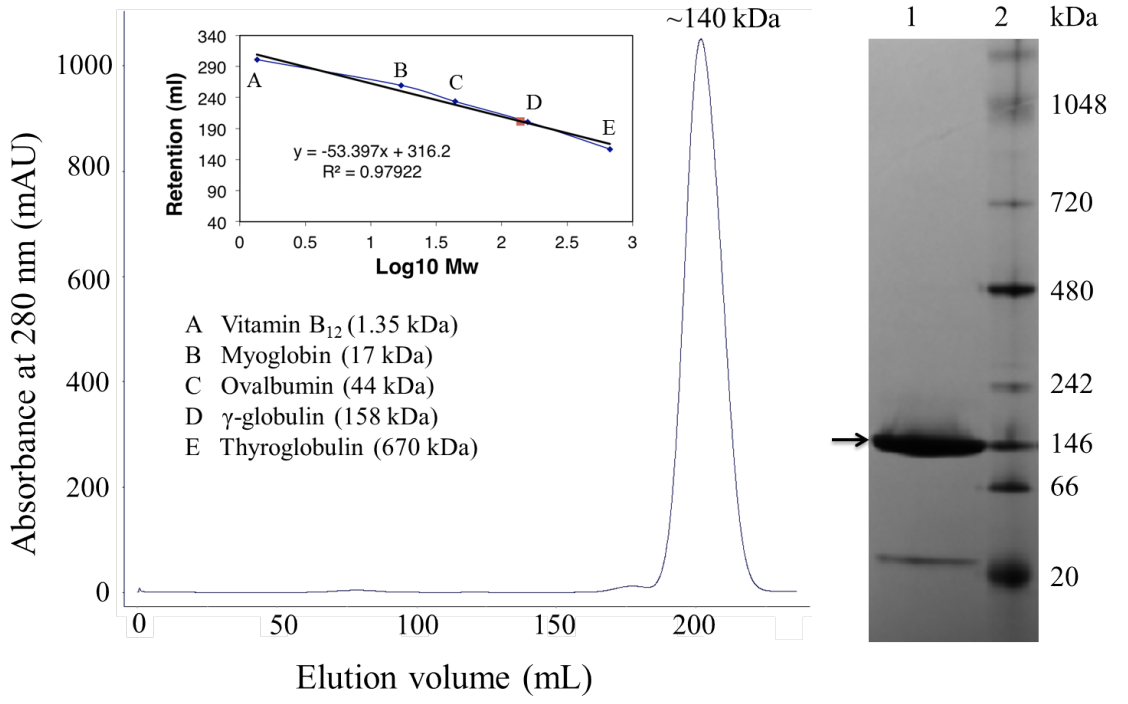

Figure 6.1 Size exclusion chromatography and native PAGE analysis.
*Bc*CCP eluted with MW of 140 kDa and displayed a 146 kDa band on native PAGE gel. The calibration curve was plotted using five standards labelled A to E. The 4-16 % Native PAGE gel shows protein (lane 1) and MW standards (lane 2).

6.4 Crystallisation and data collection

A crystallisation screen utilising commercially available formulations and was set up with a Phoenix liquid handling system (Art Robbins Instruments / Rigaku) (section 2.5.3). 100 nL^{-1} of protein solution at 10 mg mL^{-1} in 100 mM $Na(CH_3CO_2)$ pH 5.0, 150 mM NaCl and 0.5 mM $ZnSO_4$ were mixed with the same volume of reservoir and equilibrated against 70 μL^{-1} reservoir at 20 ºC. Orthorhombic rod-shaped crystals, with approximate dimensions of 30 x 10 x 10 μm, were observed after three days in conditions containing a reservoir of 0.2 M Li_2SO_4, 25 % PEG 3350 and 0.1 M Bis-Tris pH 5.5 (**Figure 6.2**). A single crystal was soaked in a cryoprotectant solution containing the reservoir buffer with 20 % glycerol prior to flash cooling at -173 ºC. Crystal was initially characterised in-house with a MicroMax-007 rotating-anode generator and R-AXIS IV^{++} dual image plate detector (Rigaku), prior to storage in liquid nitrogen. X-ray

diffraction data were subsequently collected on the ID23-2 microfocus beamline at the European Synchrotron Radiation Facility (ESRF). The small size of the crystal and its unit cell dimensions, with a particularly long *c* axis, required the use of a microfocus beam of approximately 10 microns together with a helical data collection strategy. These strategies, combined with a short exposure time of 0.6 s and a MAR mosaic 225 mm (MAR Research) charge-coupled device (CCD) detector, allowed good separation of the reflections, appropriate sensitivity and low X-ray damage, permitting a complete data set collection. Data were integrated and scaled using *MOSFLM* (Leslie, 2006) and *SCALA* (Evans, 2006).

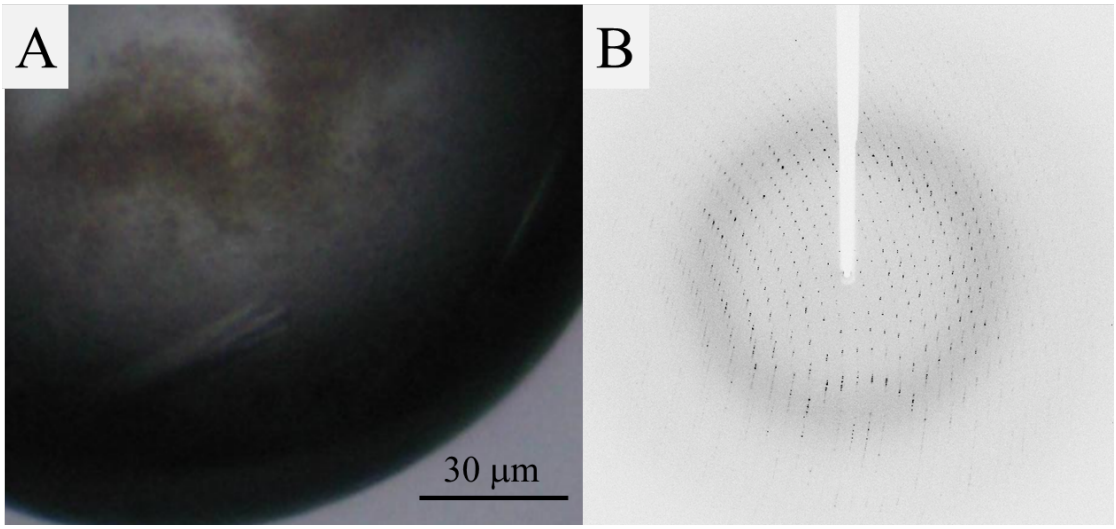

Figure 6.2 *Bc*CCP crystals.
(A) Orthorhombic rod-shaped crystals formed after three days. (B) The crystal diffracted at ESRF synchrotron using a microfocus beamline to 1.90 Å.

6.5 Structure solution and refinement

The structure was solved by molecular replacement using a poly-Ala model of the monomer of a CCP structure from *Burkholderia mallei* (*Bm*CCP), which shares 85 % sequence identity (PDB code 3k2k). Rotation and translation functions were solved using *Phaser* (McCoy *et al*., 2007). Four polypeptide chains were placed within the asymmetric unit and upon the inspection of the molecular packing in *Coot* (Emsley *et*

al., 2010) showed no steric clash. Rigid-body refinement was carried out in *REFMAC5* (Murshudov *et al.,* 2011). The resolution limits were initially set from 144.5 Å to 2.5 Å and the R_{free} was 33.3 % following rigid-body refinement. Ten cycles of restrained refinement were then performed with the resolution extended to 1.9 Å. The R_{work} and R_{free} decreased to 24.6 % and 28.6 %, respectively. The density maps for chain A were inspected in *Coot*, with some absent side chains rebuilt, and those out of density deleted. Chain A was rotated and translated to provide the other three polypeptides in the asymmetric unit and the four Zn^{2+} ions were included. At this stage, it was decided to incorporate the strategies utilised in the PDB_REDO protocols to help with efficient refinement of a sizeable model containing in excess of 1500 amino acids (Joosten *et al*., 2012). Thus, the model was subjected to several cycles of PDB_REDO refinement by Robbie Joosten (Netherlands Cancer Institute) followed again by the density maps inspection and rebuilding. Translation/ libration/ screw analysis (TLS; Painter and Merritt, 2006) was applied to create a TLS group description and the *B*-factors were reset to the Wilson *B*-factor. Following TLS refinement, the *B*-factor restraint weight was optimised with the final value of 0.30 being lower than default option. Afterwards, 30 cycles of restrained refinement with different X-ray weights were tested, with the weight 0.10 resulting in the most improved model. The model was modified to include side chains and water molecules, while Zn^{2+} ions were placed into the correct chains prior to running the second round of PDB_REDO refinement. The output model was inspected and multiple conformers were added at this point. Further rounds of restrained refinement followed, using the *B*-factor and geometric restraint weights established with PDB_REDO. Tight local non-crystallographic symmetry (NCS) restraints were imposed at the onset of the refinement and were kept until the final round of calculations. The model geometry and quality were analysed with *MolProbity* (Chen *et al.,* 2010). The crystallographic statistics are shown in **Table 6.1**.

	***Bc*CCP**
Space group	$P2_12_12_1$
Unit cell dimensions: *a, b, c* (Å)	62.90, 85.95, 289.02
Resolution[a] (Å)	40 - 1.90 (2.00 - 1.90)
No. of reflections	516512 (74589)
Unique reflections	124234 (17924)
Completeness (%)	99.9 (100.0)
Multiplicity	4.2 (4.2)
$\langle I/\sigma I \rangle$	8.4 (2.7)
Wilson *B* ($Å^2$)	13.3
Residues	
Chain A	2-156, 159-383
Chain B	2-318, 320-384
Chain C	2-156, 159-314, 320-384
Chain D	1-315, 320-384
Water / glycerol / ethylene glycol	1062 / 9 / 7
Diethylene glycol / Cl^- / Zn^{2+} / acetate	1 / 3 / 4 / 4
R_{merge}[b] (%)	13.0 (48.4)
R_{work}[c], R_{free}[d] (%)	16.4 / 20.5
Mean *B*-factors ($Å^2$)	
Protein	6.7
Ligands and waters	26.2
Ramachandran plot[e] (%)	
Most favoured	96.9
Additional allowed	2.9
Outliers	0.2
R.m.s.d. on ideal values[f]	
Bond lengths (Å) / angles (º)	0.01 / 1.14

Table 6.1 Crystallographic statistics of *Bc*CCP.
a. Values in parentheses refer to the highest resolution shell. **b.** $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$; where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements. **c.** $R_{work} = \sum_{hkl}||F_o|-|F_c||/\sum|F_o|$, where F_o is the observed structure factor and F_c is the calculated structure factor. **d.** R_{free} is the same as R_{work} except calculated with a subset, 5 %, of data that are excluded from refinement calculations. **e.** (Chen *et al*., 2010). **f.** (Engh and Huber, 1991).

6.6 *Bc*CCP structure

6.6.1 General comments

The *Bc*CCP protein crystallised in space group $P2_12_12_1$ with unit cell lengths a = 62.9Å, b = 85.9 Å, c = 289.0 Å. The asymmetric unit consists of four polypeptide chains with a V_M value of 2.17 $Å^3$ Da^{-1} and solvent content of about 45 %. Diffraction data were collected by exploiting the use of a microfocus synchrotron radiation beamline to accommodate the small size of the crystal and to help resolve the issue of a long unit

cell edge. The structure was solved by molecular replacement and then employed a slightly unconventional approach of PDB_REDO in order to carry out the efficient refinement of a relatively large asymmetric unit with data to high resolution. This procedure uses a range of automated protocols to test different models for *B*-factors, weighting restraints on geometry, optimised with respect to diffraction data parameters (Joosten *et al*., 2012). The approach was initially developed to improve the old structures held in the PDB by re-refining them using modern software.

The majority of the residues are found within well-defined electron density. The only disordered regions include the N- and C-terminal residues, residues 157-158 in chains A and C, and a number of residues in the loop (315-320) of subunits B, C and D.

A strong feature of 7 σ was observed in the difference density map in each active site near Arg226. Acetates derived from 50 mM sodium acetate present in the crystallisation conditions were modelled in. Analysis of the *Pa*CCP (Otero *et al.*, 2012; PDB code 4a39) and CP-A structures (Christianson and Lipscomb, 1986, Mangani *et al*., 1992; PDB code 1cbx) in complex with guanidinoethylmercaptosuccinic acid and benzylsuccinic acid, respectively, showed that the carboxyl group of these ligands is located at the same position as this acetate.

6.6.2 Quaternary structure

*Bc*CCP is a tetramer, forming a square-shaped structure with an opening in a centre. The tetramer is arranged as a dimer of dimers (labelled A:B and C:D in **Figure 6.3**), which is consistent with the structures of *Bm*CCP and *Shewanella denitrificans* CCP (*Sd*CCP, PDB code 3l2n). The only known exception is *Pa*CCP, reported to be monomeric (Otero *et al.*, 2012).

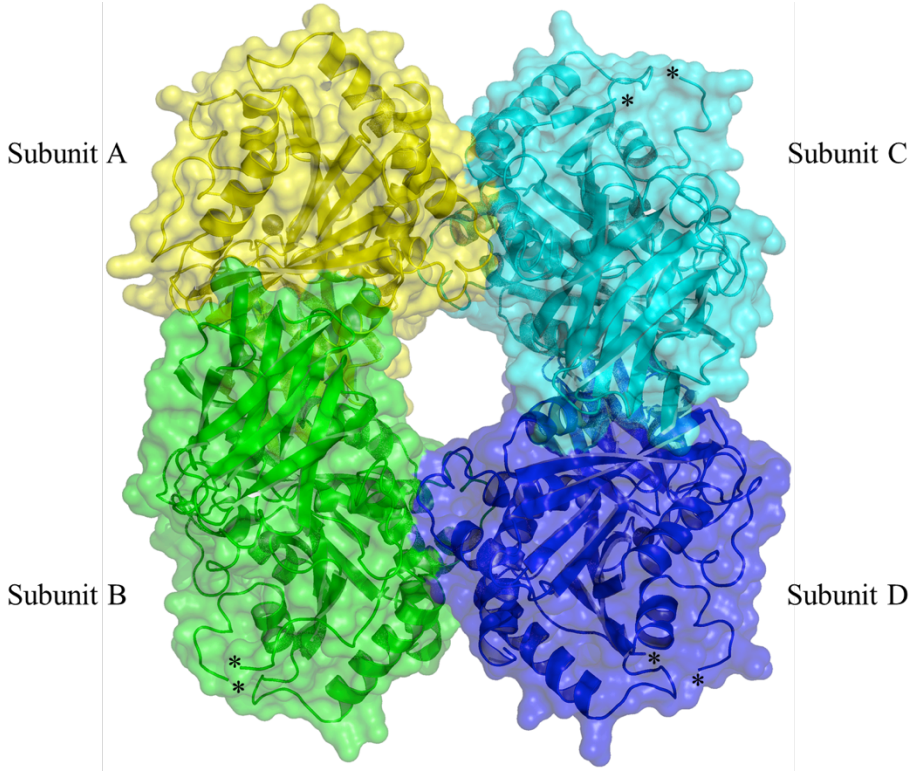

Figure 6.3 *Bc*CCP tetramer.
The structure is displayed as a semi-transparent van der Waals surface over the ribbon cartoon. Subunits are labelled A (yellow), B (green), C (cyan) and D (blue). Asterisks mark positions of disordered loops.

The average solvent accessible surface area of each *Bc*CCP subunit is approximately 15500 Å^2. The interface formed between two dimers in a tetramer (A:B and C:D) occludes 2800 Å^2, while the interface between subunits A:C (and B:D) equates to 1900 Å^2. Therefore, about 9 % of the surface area of each subunit is involved in dimer formation, with further 6 % participating in associations between a dimer of dimers. The secondary structure elements forming the interface are shown in **Figure 6.4**. Two polypeptides in a dimer are stabilised by around 24 hydrogen bonds, nine salt bridges and van der Waals interactions. Some of the residues contributing to a dimer (A:B and C:D) include Gln127 and Glu117 on α1, Asp104 on a loop between β8 and β9, Ser105 on β9, and Glu142 on a loop between β10 and β11 (not shown). There are about eight hydrogen bonds and two salt bridges between a pair of dimers in a tetramer. A few residues forming the interactions include Asp193 on a loop between α2 and α3, Lys186

on α2 and Arg201 on α3 (not shown). These residues are either conserved or substituted with equivalent in *Bm*CCP and *Sd*CCP, but they are not conserved in *Pa*CCP.

Superposition of Cα atoms of all four chains of *Bc*CCP gives an r.m.s.d in the range 0.23 – 0.35 Å, indicating that the errors associated with the structure are insignificant and subunits are essentially identical. Therefore, only subunit A is detailed, unless stated otherwise.

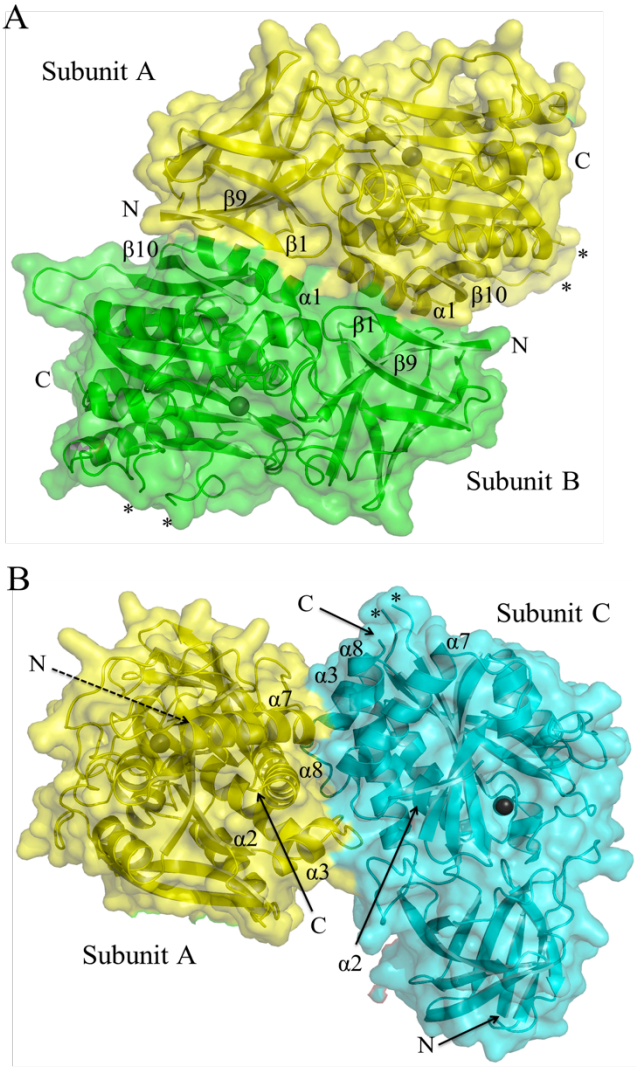

Figure 6.4 Tetramer interface.
(A) The interface between subunits in a dimer of dimers. (B) The interface between two dimers in a tetramer. The secondary structure elements lying at the interface are labelled. Disordered loops are marked with asterisks. The dashed line points to the N-terminal end, which is not visible on a cartoon.

6.6.3 Subunit structure and comparison with other CCPs

The structure of *Bc*CCP contains two domains similar to other M14 family members. The N-terminal domain consists of residues Met1 – Glu112 and the remaining 272

residues form the CP domain (**Figure 6.5**). The N-terminal domain folds into a nine-stranded antiparallel β-sandwich. Strands β2, β3, β5, β8 and β7 form one sheet and β1, β4, β9 and β6 the other. This domain of unknown function is specific for CCP proteins and absent in other carboxypeptidases. The domain might be contributing to folding, playing a regulatory role and/or might be involved in binding other proteins (Kalinina *et al.*, 2007; Rodriguez de la Vega *et al.*, 2007). The N-terminal domain is structurally conserved among CCPs, with a superposition of this domain from *Bc*CCP and *Pa*CCP giving an r.m.s.d value of 0.87 Å for over 109 Cα positions.

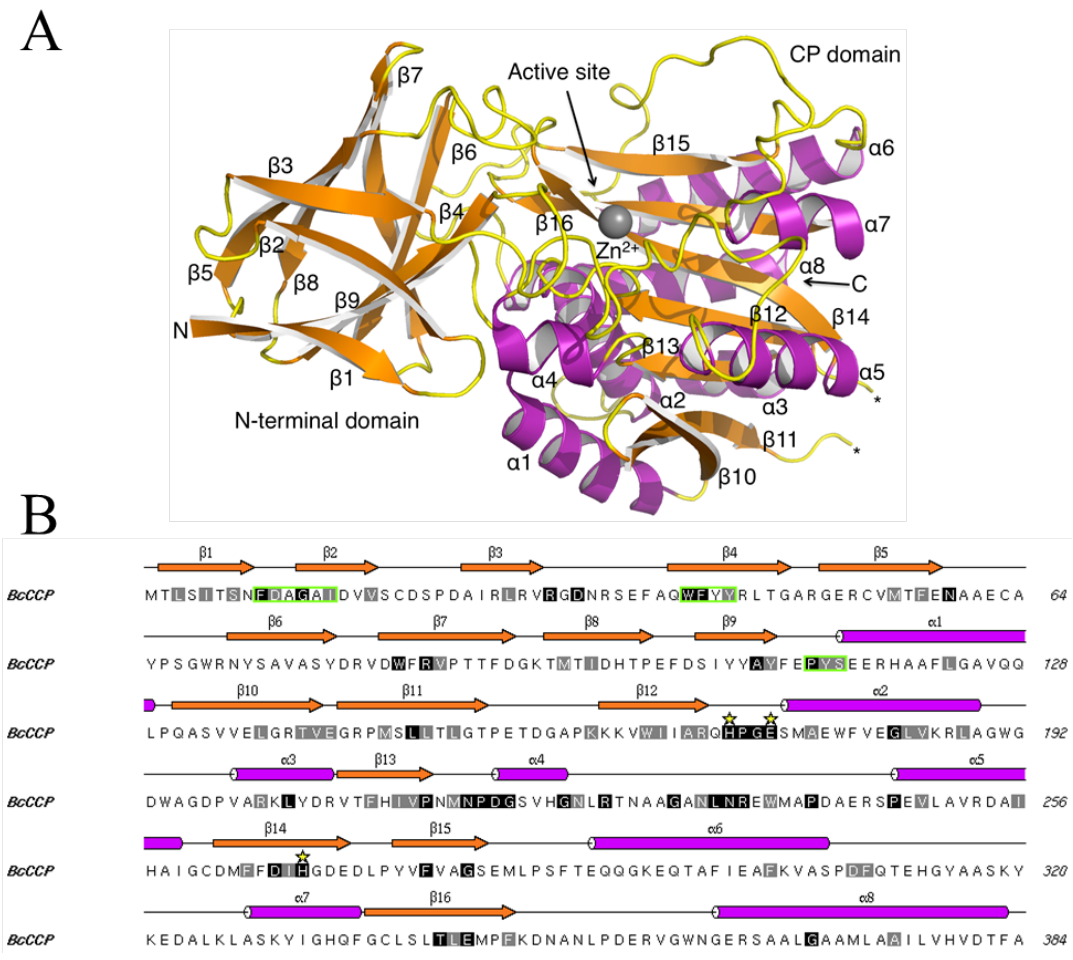

Figure 6.5 The structure of *Bc*CCP.
(A) The tertiary structure of subunit A showing α-helices (magenta) and β-strands (orange). Zn^{2+} depicted as a grey sphere. The N- and C-terminal positions are labelled, as are the α-helices and β-strands. A break in chain is marked with asterisks. (B) The primary and secondary structure of BcCCP. Residues that are strictly or highly conserved in *Bc*CCP, *Bm*CCP, *Sd*CCP, *Pa*CCP and *Ce*CCPP-6 are enclosed in black and grey, respectively. Three residues that coordinate Zn^{2+} are marked with a yellow star. The three motifs of the N-terminal domain, which are conserved in CCP members, are enclosed in green boxes.

The CP domain of *Bc*CCP consists of an α/β/α sandwich structure with a seven-stranded antiparallel β-sheet. The same structural fold is present in other M14 family carboxypeptidases such as bovine carboxypeptidase A (CP-A, **Figure 6.6**). An overlay of *Bc*CCP and bovine CP-A (PDB code 6cpa; Kim and Lipscomb, 1990), which share about 20 % sequence identity, gives an r.m.s.d of 2.03 Å for 225 Cα positions. There is a good overall fit between the central cores of both structures (**Figure 6.6**), with an exception of their flexible loops present near the active site. CP fold is highly homologous among the CCP structures, with *Bc*CCP overlaid with *Bm*CCP (PDB code 3k2k), *Sd*CCP (PDB code 3l2n) and *Pa*CCP (PDB code 4a39) giving r.m.s.d values of 0.45 Å (374 Cα atoms), 0.85 Å (362 Cα atoms), and 1.07 Å (365 Cα atoms), respectively. This structural homology correlates well with the primary sequence identity of *Bc*CCP and these proteins, which is about 85 %, 50 % and 40 %, respectively.

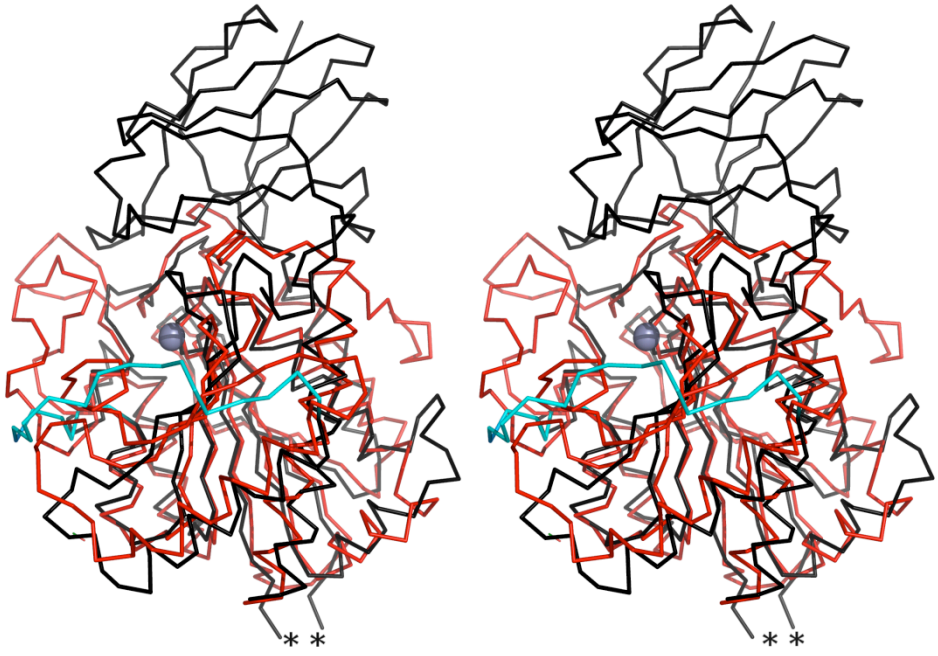

Figure 6.6 Overlay of *Bc*CCP with CP-A.
Stereo image of an overlay of Cα traces from *Bc*CCP (black) with bovine CP-A (red, PDB code 6cpa). The α6-α7 loop of *Bc*CCP is drawn in cyan. The active site Zn^{2+} is shown as a grey sphere and a loop that remains disordered in *Bc*CCP is marked with two asterisks.

*Bc*CCP contains several motifs which are conserved among the CCP family members. One of them is an Asn-Pro-Asp-Gly sequence (**Figure 6.5B**), which likely helps with domain folding and contributes partially to the formation of the active site (**Figure 6.7**). In this motif, Asn216, located between α4 and β13, forms hydrogen bonds with the main chain Gln170 and a conserved Gly219. This in turn links the β12 and α4 together. In addition, Asp218 establishes hydrogen bonds to the side chain of Arg144 and the main chain of Glu142, both residues located on a loop between β10 and β11. This chain of hydrogen-bonding forms part of a larger network of interactions that hold β12 as well as β14 (not shown) in place, thereby contributing to formation of the active site because these β-strands provide His171 and His268, respectively to coordinate the catalytic Zn^{2+}.

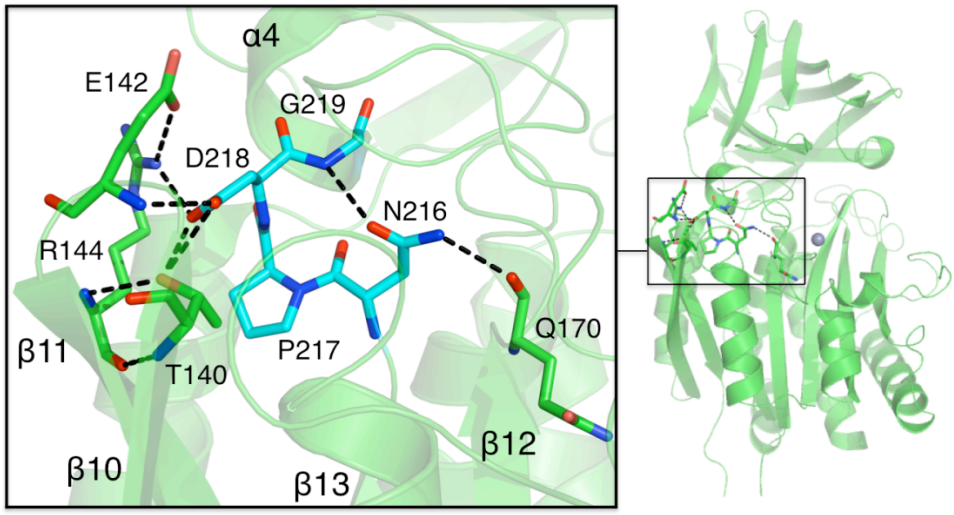

Figure 6.7 The conserved Asn-Pro-Asp-Gly motif.
On the left hand side, in stick representation, are shown the residues and selected hydrogen bonding interactions, as dashed black lines. The Asn-Pro-Asp-Gly residues are drawn in cyan. The right panel is a ribbon diagram of a subunit to indicate the position of this motif with respect to the overall fold.

CCP proteins vary in length, with the mammalian proteins being considerably larger than their bacterial homologues due to the N- and C-terminal extensions plus numerous insertions. For example, CCP1 is the largest mammalian isoform with around 1200 residues, of which 700 residues are the N-terminal extension. CCP6 is the shortest

mammalian isoform as it lacks the usually long N- and C-terminal extensions. Thus, the extra segments vary considerably in length, but their role is currently unknown. However, all CCP members contain three short sequence motifs within their N-terminal domain, which are absent in other carboxypeptidase families (Kalinina *et al.*, 2007). The structure reveals that these conserved motifs are located in close proximity, even though they are distant in the primary amino acid sequence (**Figure 6.5B**; **6.8**). Motif I F[D,E]xGx[L,I] (residues 9-14) is located at the N-terminus, followed by motif II W[F,Y][Y,N][F,Y] (residues 40-43) 60 residues downstream. This motif is located at the centre of the N-terminal domain. The conserved Trp40 and Phe41 form a hydrophobic core with other large hydrophobic residues such as Phe9, Phe57, Trp69, Trp83, Tyr110 and Phe111. Final motif III P[F,Y][S,T] (residues 113-115) is found at the N-terminus of α1 on the surface of the domain. This segment forms part of interface between the N-terminal and the CP domains. These motifs are also present in *Bm*CCP, *Pa*CCP and *Sd*CCP structures (**Figure 6.5B**), and support the notion that the three conserved motifs are important for correct protein folding. However, a possibility that they might have additional roles cannot be eliminated.

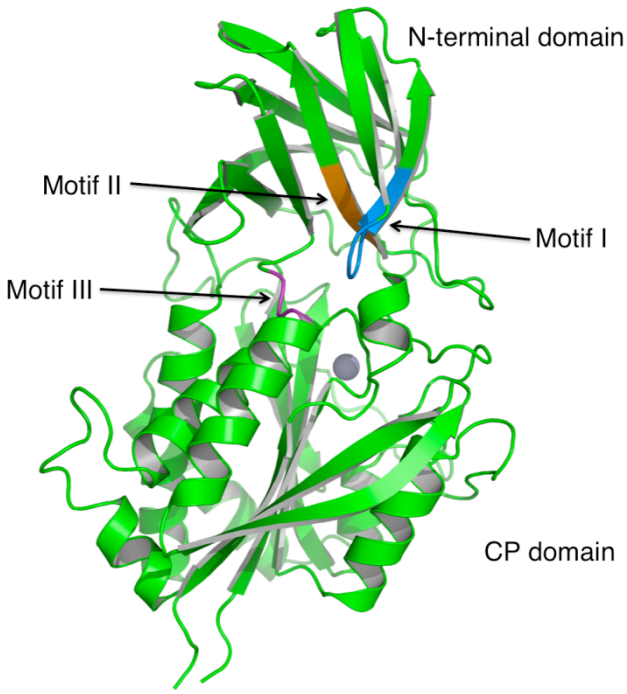

Figure 6.8 Three motifs common to the CCP family.
These are: motif I (marine), motif II (orange), and motif III (purple). The active site Zn^{2+} is a grey sphere and the three-conserved motifs of the N-terminal domain found in all M14D family members are shown in marine and are labeled accordingly.

6.6.4 The active site and specificity

In *Bc*CCP, the active site is located in a groove, the floor of which is formed predominantly by residues in strands β14, β15 and β16. The walls of the groove are made out of four loops (α4-α5, α6-α7, β3-β4 and β5-β6) and the β14- β15 turn (**Figure 6.5**). Here, the catalytic Zn^{2+} is penta-coordinated in a distorted tetrahedral geometry by His171, Glu174, His268 and a water molecule (**Figure 6.9A**). In addition, as mentioned earlier the active site also contains an acetate derived from the crystallisation conditions, which is held in place by Arg226, Arg236 and Asn235 (**Figure 6.9A**).

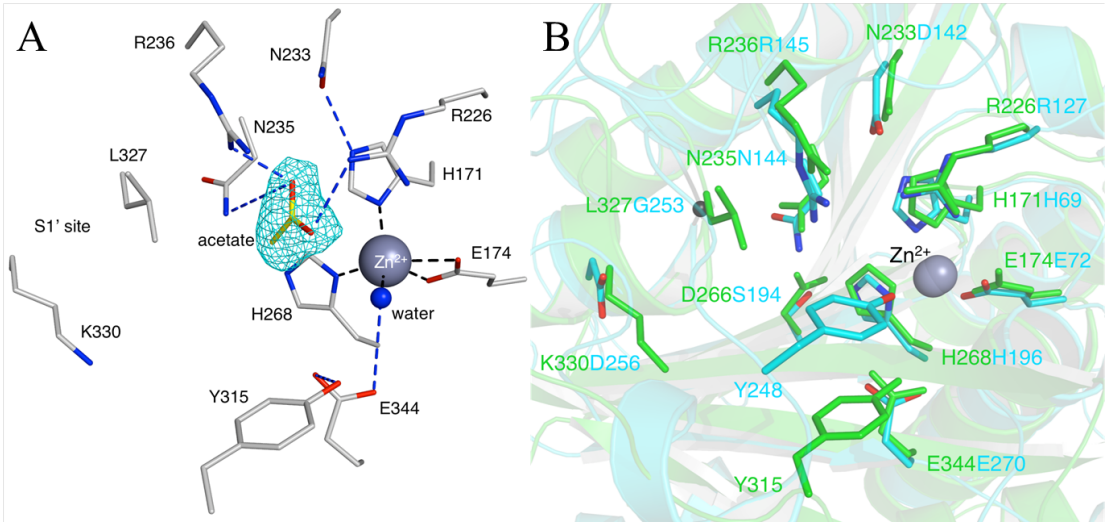

Figure 6.9 The active site of *Bc*CCP.
(A) Zn^{2+} coordination in *Bc*CCP. Selected residues in the active site and acetate are displayed and coloured as C white, or yellow for acetate, O red, N blue. The $2F_o$-F_c electron density map is shown for acetate, where F_o are the observed and F_c are the calculated structure factors. The map is contoured at 1σ (cyan chicken wire). Zn^{2+} and a water ligand are shown as spheres, coloured grey and marine respectively. Zinc coordination is shown with black dashed lines and blue dashed lines represent hydrogen bonding interactions. (B) A comparison of *Bc*CCP and bovine CP-A (PDB code 6cpa; Kim and Lipscomb, 1990) active sites. Residue side chains are depicted as sticks with atomic positions coloured O red, N blue and C cyan for bovine CP-A or all green for *Bc*CCP. The Cα of Gly253 CP-A is a small black sphere. Zn^{2+} ions are overlapping grey spheres.

The bacterial CCP active sites are structurally highly conserved and this conservation extends to cover many other carboxypeptidases. This is also true for the active sites of *Bc*CCP and bovine CPA (**Figure 6.9B**), even though the two share only 21 % overall sequence identity. His171 and Glu174 residues are part of the His-Pro-Gly-Glu motif,

which is highly conserved in all CCP proteins (**Figure 6.5B**). In *Bc*CCP and *Bm*CCP, the nearby Arg169 helps to stabilise the spatial conformation of this motif by making a hydrogen bond to the main chain carbonyl of Pro172. Similar stabilising contacts are also observed in other available structures, but they involve different residues.

In the case of bovine CPA, in addition to zinc coordinating ligands, other important residues involved in catalysis and possibly in substrate binding are Glu270 and Arg127. Arg127 helps to stabilize the oxyanion hole by binding to the carbonyl bond, while Glu270 acts as the general base for catalysis (Kim and Lipscomb, 1990). In the *Bc*CCP structure, the residues in the equivalent positions are Glu344 and Arg226 (**Figure 6.9B**). Other residues conserved in the active site throughout the M14 family include Asn144 and Arg145 of bovine CP-A, or equivalent Arg236 and Asn235 in *Bc*CCP. In the case of CP-A, these two residues interact with the C-terminal carboxylate group of the substrate. The Tyr248 in CPA structure, whose involvement in the catalytic process is still uncertain corresponds to Tyr315 in *Bc*CCP (Rees *et al*., 1981; Gardell *et al.*, 1985; Cho *et al.*, 2001). It is worth noting that in addition to Tyr315, *Bc*CCP contains another tyrosine residue Tyr320 near its active site. Both of these are located on an α6-α7 loop. Tyr320 is located on a more flexible part, which is found in two different orientations in *Bc*CCP and *Bm*CCP structures (**Figure 6.10**). The *Bc*CCP structure displays an open configuration, whereas the *Bm*CCP structure contains a closed configuration with Tyr320 directed into the active site. However, many loop residues including Tyr320 are not conserved among other CCP members.

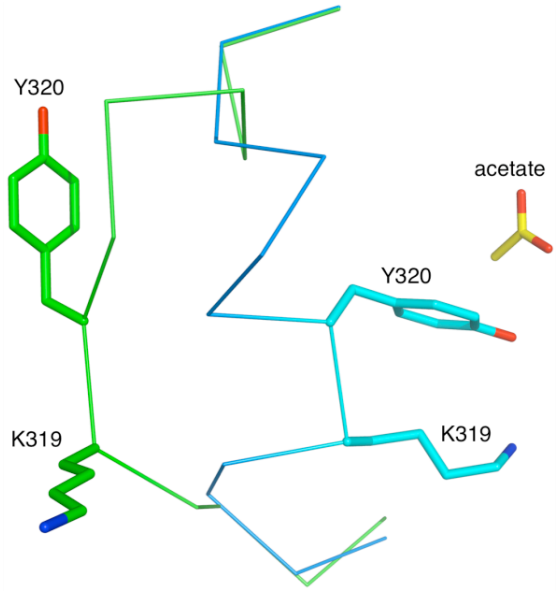

Figure 6.10 The flexible segment of the α6-α7 loop.
The *Bc*CCP and *Bm*CCP (PDB code 3k2k) subunit structures are overlaid and the Cα trace for residues 315 to 326 shown coloured green and blue/cyan respectively. The side chains of Lys319 and Tyr320 are shown as sticks.

In bovine CP-A and rat CP-A2, the substrate specificity is mostly determined by only a few residues placed in the vicinity of the S1' site of the enzyme that binds the carboxyl terminus of the substrate (Faming *et al.,* 1991). In the case of CP-A, these residues are Ile243, Asn144, Ala250, Gly253, Ile255, Asp256 and Thr268 (Garcia-Saez *et al*., 1997). The residues Ile243, Ala250 and Tyr248 are found on a flexible loop, capping the S1' site. Asn144 and Thr268 are both conserved in all CCP proteins and correspond to Asn235 and Thr342 in *Bc*CCP. Ile255 in CP-A corresponds to Ser329 in *Bc*CCP that is based at the bottom of the S1' pocket. The nearby located Gly253 and Asp256 in CP-A are replaced with Leu327 and Lys330 (**Figure 6.9B**), making the S1' site in *Bc*CCP shallower. The Lys330 is also present in *Bm*CCP and might confer the specificity of the *Burkholderia* enzymes to act as the deglutamylases. Asn321 is identically placed in this position in *Sd*CCP and *Pa*CCP, thereby their S1' subsites would still be able to accommodate the P1 glutamate on the substrate protein.

In vitro assays were performed using a furylacryloyl glutamate derivative to test *Bc*CCP for deglutamylase activity (data not shown). However, the hydrolysis reaction was inefficient, suggesting that either BcCCP might not possess deglutamylase activity, might require its specific substrate or might need activation similarly to bovine CP-A.

6.7 Conclusion

The structure of a zinc-dependent *Bc*CCP has been determined to 1.90 Å resolution. The crystals produced were small and when first analysed revealed a relatively long unit cell *c* dimension. For these reasons, the diffraction data were collected using a synchrotron microfocus beamline, which allowed the acquisition of a full data set. *Bc*CCP contains relatively large asymmetric unit with four polypeptides, in total about 1500 residues. An unconventional strategy was chosen to speed up the refinement process, which utilised the automated protocols of PDB_REDO coupled with manual model-map inspections and rebuilding. This method led to an efficient refinement, highlighting the power of such protocols. The crystal structure revealed that both domains of *Bc*CCP are structurally homologous to other solved structures of bacterial CCPs. The N-terminal domain of unknown function was to date only identified in CCP proteins, whereas the C-terminal domain is conserved in the M14 family. Eukaryotic CCPs were reported to be able to act as deglutamylases, and *Bc*CCP might have the same function. The presence of Lys330 near the S1' site provides further support to this. Nonetheless, deglutamylase activity was not observed against a furylacryloyl glutamate derivative. This indicates *Bc*CCP might not be a deglutamylase; it might need activation or catalyse selective substrates, which to date were not identified. Therefore, further studies are required to answer these questions.

References

Abada R, Dreyfuss-Grossman T, Herman-Bachinsky Y, Geva H, Masa SR and Sarid R (2008) SIAH-1 interacts with the Kaposi's sarcoma-associated herpesvirus-encoded ORF45 protein and promotes its ubiquitylation and proteasomal degradation. *J Virol* **82**: 2230-2240

Aravind L and Koonin EV (2000) SAP – a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* **25**: 112-114

Ban R, Matsuzaki H, Akashi T, Sakashita G, Taniquichi H, Park SY, Tanaka H, Furukawa K and Urano T (2009) Mitotic regulation of the stability of microtubule plus-end tracking protein EB3 by ubiquitin ligase SIAH-1 and Aurora mitotic kinases. *J Biol Chem* **284**: 28367-28381

Battye TGG, Kontogiannis L, Johnson O, Powell HR and Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst* D**67**: 271-281

Bawa-Khalfe T, Cheng J, Wang Z and Yeh ET (2007) Induction of the SUMO-specific protease 1 transcription by the androgen receptor in prostate cancer cells. *J Biol Chem* **282**: 37341-37349

Bayer P, Arndt A, Metzger S, Mahajan R, Melchior F, Jaenicke R and Becker J (1998) Structure determination of the small ubiquitin-related modifier SUMO-1. *J Mol Biol* **280**: 275-286

Bembenek SD, Tounge BA and Reynolds CH (2009) Ligand efficiency and fragment-based drug discovery. *Drug Discov Today* **14**: 278-283

Bhattacharya S, Lee YT, Michowski W, Jastrzebska B, Filipek A, Kuznicki J and Chazin WJ (2005) The modular structure of SIP facilitates its role in stabilizing multiprotein assemblies. *Biochemistry* **44**: 9462-9471

Boehm J, He Y, Greiner A, Staudt L and Wirth T (2001) Regulation of BOB.1/OBF.1 stability by SIAH. *EMBO J* **20**: 4153-4162

Bogan AA and Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**: 1-9

Bohacek RS, McMartin C and Guida WC (1996) The art and practice of strycture-based drug design: a molecular modelling perspective. *Med Res Rev* **16**: 3-50

Bond CS and Schüttelkopf AW (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Cryst* D**65**: 510-512

Boulton SJ, Brook A, Staehling-Hampton K, Heitzler P and Dyson N (2000) A role for Ebi in neuronal cell cycle control. *EMBO J* **19**: 5376-5386

Brauckhoff A, Malz M, Tschaharganeh D, Malek N, Weber A, Riener MO, Soll C, Samarin J, Bissinger M, Schmidt J, Longerich T, Ehemann V, Schirmacher P and Breuhahn K (2011) Nuclear expression of the ubiquitin ligase seven in absentia homolog (SIAH)-1 induces proliferation and migration of liver cancer cells. *J Hepatol* **55**: 1049-1057

Brough PA, Barril X, Borgognoni J, Chene P, Davies NG, Davis B, Drysdale MJ, Dymock B, Eccles SA, Garcia-Echeverria C, Fromont C, Hayes A, Hubbard RE, Jordan AM, Jensen MR, Massey A, Merrett A, Padfield A, Parsons R, Radimerski T, Raynaud FI, Robertson A, Roughley SD, Schoepfer J, Simmonite H, Sharp SY, Surgenor A, Valenti M, Walls S, Webb P, Wood M, Workman P and Wright L (2009) Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. *J Med Chem* **52**: 4794-4809

Bruzzoni-Giovanelli H, Fernandez P, Veiga L, Podgorniak MP, Powell DJ, Candeias M. M, Mourah S, Calvo F and Marín M (2010) Distinct expression patterns of the E3 ligase SIAH-1 and its partner Kid/KIF22 in normal tissues and in the breast tumoral processes. *J Exp Clin Cancer Res* **29**: 10

Buchwald M, Pietschmann K, Brand P, Günther A, Mahajan NP, Heinzel T and Krämer OH (2012) SIAH ubiquitin ligases target the nonreceptor tyrosine kinase ACK1 for ubiquitinylation and proteasomal degradation. *Oncogene* Epub ahead of print

Budhidarmo R, Nakatani Y and Day CL (2012) RINGs hold the key to ubiquitin transfer. *Trends Biochem Sci* **37**: 58-65

Bursen A, Moritz S, Gaussmann A, Moritz S, Dingermann T and Marschalek R (2004) Interaction of AF4 wild-type and AF4.MLL fusion protein with SIAH proteins: indication for t(4;11) pathobiology? *Oncogene* **23**: 6237-6249

Carthew RW and Rubin GM (1990) Seven in absentia, a gene required for specification of R7 cell fate in the *Drosophila* eye. *Cell* **63**: 561-577

Carthew RW, Neufeld TP and Rubin GM (1994) Identification of genes that interact with the sina gene in *Drosophila* eye development. *Proc Natl Acad Sci USA* **91**: 11689-11693

Chen IJ and Hubbard RE (2009) Lessons for fragment library design: analysis of output from multiple screening campaigns. *J Comput Aided Mol Res* **23**: 603-620

Chen VB, Arendal WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS and Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst* D**66**: 12-21

Chen Y, Wen D, Huang Z, Huang M, Luo Y, Liu B, Lu H, Wu Y, Peng Y and Zhang J (2012) 2-(4-Chlorophenyl)-2-oxoethyl 4-benzamidobenzoate derivatives, a novel class of SENP1 inhibitors: Virtual screening, synthesis and biological evaluation. *Bioorg Med Chem Lett* **22**: 6867-6870

Cheng CH, Lo YH, Liang SS, Ti SC, Lin FM, Yeh CH, Huang HY and Wang TF (2006) SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of *Saccharomyces cerevisiae*. *Genes Dev* **20**: 2067-2081

Cho JH, Kim DH, Lee KJ, Kim DH and Choi KY (2001) The role of Tyr248 probed by mutant bovine carboxypeptidase A: insight into the catalytic mechanism of carboxypeptidase A. *Biochemistry* **40**: 10197-10203

Christianson DW and Lipscomb WN (1986) X-ray crystallographic investigation of substrate binding to carboxypeptidase A at subzero temperature. *Proc Natl Acad Sci USA* **83**: 7568-7572

Christianson DW and Lipscomb WN (1989) Carboxypeptidase A. *Acc Chem Res* **22**: 62-69

Confalonieri S, Quarto M, Goisis G, Nuciforo P, Donzelli M, Jodice G, Pelosi G, Viale G, Pece S and Di Fiore PP (2009) Alterations of ubiquitin ligases in human cancer and their association with the natural history of the tumor. *Oncogene* **28**: 2959-2968

Congreve M, Carr R, Murray C and Jhoti H (2003) A 'rule of three' for fragment-based lead discovery? *Drug Discov Today* **8**: 876-877

Cooper SE, Murawsky CM, Lowe N and Travers AA (2008) The modes of degradation of the tramtrack transcription factors by Siah homologues. *J Biol Chem* **283**: 1076-1083

Crone J, Glas C, Schultheiss K, Moehlenbrink J, Krieghoff-Henning E and Hofmann TG (2011) Zyxin is a critical regulator of the apoptotic HIPK2-p53 signaling axis. *Cancer Res* **71**: 2350-2359

Cummings MD, Farnum MA and Nelen MI (2006) Universal screening methods and applications of ThermoFluor. *J Biomol Screen* **11**: 854-863

Dalvit C, Pevarello P, Tatò M, Veronesi M, Vulpetti A and Sundström M (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J Biochem NMR* **18**: 65-68

Della NG, Senior PV and Bowtell DD (1993) Isolation and characterisation of murine homologues of the *Drosophila* seven in absentia gene (sina). *Development* **117**: 1333-1343

Delmar M, Coombs W, Sorgen P, Duffy HS and Taffet SM (2004) Structural bases for the chemical regulation of Connexin43 channels. *Cardiovasc Res* **62**: 268-275

Deshaies RJ and Joazeiro CA (2009) RING domain E3 ubiquitin ligases. *Annu Rev Biochem* **78**: 399-434

Dimitrova YN, Li J, Lee YT, Rios-Esteves J, Friedman DB, Choi HJ, Weis WI, Wang CY and Chazin WJ (2010) Direct ubiquitination of β-catenin by Siah-1 and regulation by the exchange factor TBL1. *J Biol Chem* **285**: 13507-13516

Dou H, Buetow L, Hock A, Sibbet GJ, Vousden KH and Huang DT (2012) Structural basis for autoinhibition and phosphorylation-dependent activation of c-Cbl. *Nat Struct Mol Biol* **19**: 184-192

Edde B, Rossier J, Le Caer JP, Desbruyeres E, Gros F and Denoulet P (1990) Posttranslational glutamylation of α-tubulin. *Science* **247**: 83-85

Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Cryst* D**66**: 486-501

Engh RA and Huber R (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst* A**47**: 392-400

Ertl P (2003) Cheminformatics analysis of organic substituents: Identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J Chem Inf Comput Sci* **43**: 374-380

Espinas ML, Jimenez-Garcia E, Vaquero A, Canudas S, Bernues J and Azorin F (1999) The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. *J Biol Chem* **274**: 16461-16469

Evans P (2006) Scaling and assessment of data quality. *Acta Cryst* D**62**: 72-82

Faming Z, Kobem B, Stewart CB, Rutter WJ and Goldsmith EJ (1991) Structural evolution of an enzyme specificity. The structure of rat carboxypeptidase A2 at 1.9- Å resolution. *J Biol Chem* **266**: 24606-24612

Fanelli M, Fantozzi A, De Luca P, Caprodossi S, Matsuzawa S, Lazar MA, Pelicci PG and Minucci S (2004) The coiled-coil domain is the structural determinant for mammalian homologues of *Drosophila* Sina-mediated degradation of promyelocytic leukemia protein and other tripartite motif proteins by the proteasome. *J Biol Chem* **279**: 5374-5379

Freemont PS, Hanson IM and Trowsdale J (1991) A novel cysteine-rich sequence motif. *Cell* **64**: 483-484

Fujita K, Horikawa I, Mondal AM, Jenkins LM, Appella E, Vojtesek B, Bourdon JC, Lane DP and Harris CC (2010) Positive feedback between p53 and TRF2 during telomere-damage signalling and cellular senescence. *Nat Cell Biol* **12**: 1205-1212

Fukuba H, Takahashi T, Jin HG, Kohriyama T and Matsumoto M (2008) Abundance of aspargynyl-hydroxylase FIH is regulated by Siah-1 under normoxic conditions. *Neurosci Lett* **433**: 209-214

García-Sáez I, Reverter D, Vendrell J, Avilés FX and Coll M (1997) The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *EMBO J* **16**: 6906-6913

Gardell SJ, Craik CS, Hilvert D, Urdea MS and Rutter WJ (1985) Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase A does not play a crucial role in catalysis. *Nature* **317**: 551-555

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A (2005) (In) Walker JM (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607

Gaston BM, Carver J, Doctor A and Palmer LA (2003) S-nitrosylation signalling in cell biology. *Mol Interv* **3**: 253-263

GE Healthcare (2008) Biacore Sensor Surface Handbook BR-1005-71 Edition AB

Germani A, Bruzzoni-Giovanelli H, Fellous A, Gisselbrecht S, Varin-Blank N and Calvo F (2000) SIAH-1 interacts with α-tubulin and degrades the kinesin Kid by the proteasome pathway during mitosis. *Oncogene* **19**: 5997-6006

Germani A, Prabel A, Mourah S, Podgorniak MP, Di Carlo A, Ehrlich R, Gisselbrecht S, Varin-Blank N, Calvo F and Bruzzoni-Giovanelli H (2003) SIAH-1 interacts with CtIP and promotes its degradation by the proteasome pathway. *Oncogene* **22**: 8845-8851

Giannetti AM (2011) From experimental design to validated hits a comprehensive walk-through of fragment leas identification using surface plasmon resonance. *Methods Enzymol* **493**: 169-218

Glozak MA, Sengupta N, Zhang X and Seto E (2005) Acetylation and deacetylation of non-histone proteins. *Gene* **363**: 15-23

Gomis-Rüth FX (2008) Structure and mechanism of Metallocarboxypeptidases. *Crit Rev Biochem Mol Biol* **43**: 319-345

Gong L, Millas S, Maul GG and Yeh ET (2000) Differential regulation of sentrinized proteins by a novel sentrin-specific protease. *J Biol Chem* **275**: 3355-3359

Grabbe C, Husnjak K and Dikic I (2011) The special and temporal organization of ubiquitin networks. *Nat Rev Mol Cell Biol* **12**: 295-307

Grewal SI and Rice JC (2004) Regulation of heterochromatin by histone methylation and small RNAs. *Curr Opin Cell Biol* **16**: 230-238

Grishina I, Debus K, García-Limones C, Schneider C, Shresta A, García C, Calzado MA and Schmitz ML (2012) SIAH-mediated ubiquitination and degradation of acetyl-transferases regulate the p53 response and protein acetylation. *Biochim Biophys Acta* **1823**: 2287-2296

Habelhah H, Laine A, Erdjument-Bromage H, Tempst P, Gershwin ME, Bowtell DD and Ronai Z (2004) Regulation of 2-oxoglutarate (α-ketoglutarate) dehydrogenase stability by the RING finger ubiquitin ligase Siah. *J Biol Chem* **279**: 53782-53788

Hajduk PJ, Meadows RP and Fesik SW (1997) Discovering high-affinity ligands for proteins. *Science* **278**: 497-499

Halazonetis TD, Georgopoulos K, Greenberg ME and Leder P (1988) c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. *Cell* **55:** 917-924

Hämäläinen MD, Zhukov A, Ivarsson M, Fex T, Gottfries J, Karlsson R and Björsne M (2008) Label-free primary screening and affinity ranking of fragment libraries using parallel analysis of protein panels. *J Biomol Screen* **13**: 202-209

Hann MM, Leach AR and Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci* **41**: 856-864

Hara MR and Snyder SH (2006) Nitric oxide-GAPDH-Siah: a novel cell death cascade. *Cell Mol Neurobiol* **26**: 527-538

Hattersley N, Shen L, Jaffray EG and Hay RT (2011) The SUMO protease SENP6 is a direct regulator of PML nuclear bodies. *Mol Biol Cell* **22**: 78-90

He HT, Fokas E, You A, Engenhart-Cabillic R and An HX (2010) Siah1 proteins enhance radiosensitivity of human breast cancer cells. *BMC Cancer* **10**: 403

Hershko A and Ciechanover A (1998) The ubiquitin system. *Annu Rev Biochem* **67**: 425-479

Hickey CM, Wilson NR and Hochstrasser M (2012) Function and regulation of SUMO proteases. *Nat Rev Mol Cell Biol* **13**: 755-766

Hoeller D and Dikic I (2009) Targeting the ubiquitin system in cancer therapy. *Nature* **458**: 438-444

Holdgate GA, Anderson M, Edfeldt F and Geschwindner S (2010) Affinity-based, biophysical methods to detect and analyse ligand binding to recombinant proteins: matching high information content with high throughput. *J Struct Biol* **172**: 142-157

Holloway AJ, Della NG, Fletcher CF, Largespada DA, Copeland NG, Jenkins NA and Bowtell DD (1997) Chromosomal mapping of five highly conserved murine homologues of the *Drosophila* RING finger gene seven-in-absentia. *Genomics* **41**: 160-168

Holm L and Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**: W545-W549

Hopkins AL, Groom CR and Alex A (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov Today* **9**: 430-431

Hourdou ML, Guinand M, Vacheron MJ, Michel G, Denoroy L, Duez C, Englebert S, Joris B, Weber G and Ghuysen JM (1993) Characterization of the sporulation-related gamma-D-glutamyl-(L)meso-dianopimelic-acid-hydrolysing peptidase I of *Bacillus sphaericus* NCTC 9602 as a member of the metallo(zinc) carboxypeptidase A family. Modular design of the protein. *Biochem J* **292**: 563-570

House CM, Frew IJ, Huang HL, Wiche G, Traficante N, Nice E, Catimel B and Bowtell DD (2003) A binding motif for Siah ubiquitin ligase. *Proc Natl Acad Sci USA* **100**: 3101-3106

House CM, Hancock NC, Möller A, Cromer BA, Fedorov V, Bowtell DD, Parker MW and Polekhina G (2006) Elucidation of the substrate binding site of Siah ubiquitin ligase. *Structure* **14**: 695-701

House CM, Möller A and Bowtell DD (2009) Siah proteins: novel drug targets in the Ras and hypoxia pathways. *Cancer Res* **69**: 8835-8838

Hu G and Fearon ER (1999) Siah-1 N-terminal RING domain is required for proteolysis function, and C-terminal sequences regulate oligomerisation and binding to target proteins. *Mol Cell Biol* **19**: 724-732

Hu G, Zhang S, Vidal M, Baer JL, Xu T and Fearon ER (1997) Mammalian homologs of seven in absentia regulate DCC via the ubiquitin-proteasome pathway. *Genes Dev* **11**: 2701-2714

Huang L, Kinnucan E, Wang G, Beaudenon S, Howley PM, Huibregtse JM and Pavletich NP (1999) Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **286**: 1321-1326

Hubbard RE (2008) Fragment approaches in structure-based drug discovery. *J Synchrotron Radiat* **15**: 227-230

Hubbard RE (2011) Structure-based drug discovery and protein targets in the CNS. *Neuropharmacology* **60**: 7-23

Hubbard RE and Murray JB (2011) Experiences in fragment-based lead discovery. *Methods Enzymol* **493**: 509-531

Ishikawa K, Nash SR, Nishimune A, Neki A, Kaneko S and Nakanishi S (1999) Competitive interaction of seven in absentia homolog-1A and Ca^{2+}/calmodulin with the cytoplasmic tail of group 1 metabotropic glutamate receptors. *Genes Cells* **4**: 381-390

Ivanov D, Stone JR, Maki JL, Collins T and Wagner G (2005) Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain. *Mol Cell* **17**: 137-143

Jackson SP and Durocher D (2013) Regulation of DNA damage responses by ubiquitin and SUMO. *Mol Cell* **49**: 795-807

Jesenberger V and Jentsch S (2002) Deadly encounter: ubiquitin meets apoptosis. *Nat Rev Mol Cell Biol* **3**: 112-121

Jhoti H (2005) A new school for screening. *Nat Biotechnol* **23**: 184-186

Jiang X, Yu Y, Yang HW, Agar NY, Frado L and Johnson MD (2010) The imprinted gene PEG3 inhibits Wnt signalling and regulates glioma growth. *J Biol Chem* **285**: 8472-8480

Johnsen SA, Subramaniam M, Monroe DG, Janknecht R and Spelsberg TC (2002) Modulation of transforming growth factor β (TGFβ)/Smad transcriptional responses through targeted degradation of TGFβ-inducible early gene-1 by human seven in absentia homologue. *J Biol Chem* **277**: 30754-30759

Johnson ES (2004) Protein modification by SUMO. *Annu Rev Biochem* **73**: 355-382

Joosten RP, Joosten K, Murshudov GN and Perrakis A (2012) PDB_REDO: constructive validation, more than just looking for errors. *Acta Cryst* D**68**: 484-496

Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Cryst* A**32**: 922-923

Kaikkonen S, Jääskeläinen T, Karvonen U, Rytinki MM, Makkonen H, Gioeli D, Paschal BM and Palvimo JJ (2009) SUMO-specific protease 1 (SENP1) reverses the hormone-augmented SUMOylation of androgen receptor and modulates gene responses in prostate cancer cells. *Mol Endocrinol* **23**: 292-307

Kalinina E, Biswas R, Berezniuk I, Hermoso A, Aviles FX and Fricker LD (2007) A novel subfamily of mouse cytosolic carboxypeptidases. *FASEB J* **21**: 836-850

Kelley LA and Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**: 363-371

Kim H and Lipscomb WN (1990) Crystal structure of the complex of carboxypeptidase A with a strongly bound phosphonate in a new crystalline form: comparison with structures of other complexes. *Biochemistry* **29**: 5546-5555

Kim H, Jeong W, Ahn K, Ahn C and Kang S (2004) Siah-1 interacts with the intracellular region of polycystin-1 and affects its stability via the ubiquitin-proteasome pathway. *J Am Soc Nephrol* **15**: 2042-2049

Kimura Y, Kurabe N, Ikegami K, Tsutsumi K, Konishi Y, Kaplan OI, Kunitomo H, Iino Y, Blacque OE and Setou M (2010) Identification of tubulin deglutamylase among *Caenorhabditis elegans* and mammalian cytosolic carboxypeptidases (CCPs). *J Biol Chem* **285**: 22936-22941

Kirkin V and Dikic I (2011) Ubiquitin networks in cancer. *Curr Opin Genet Dev* **21**: 21-28

Komander D and Rape M (2012) The ubiquitin code. *Annu Rev Biochem* **81**: 203-229

Krämer OH, Stauber RH, Bug G, Hartkamp J and Knauer SK (2013) SIAH proteins: critical roles in leukemogenesis. *Leukemia* **27**: 792-802

Krissinel E and Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst* D**60**: 2256-2268

Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**: 774-797

Kroetz MB (2005) SUMO: a ubiquitin-like protein modifier. *Yale J Biol Med* **78**: 197-201

Leach AR and Hann MM (2011) Molecular complexity and fragment-based drug discovery: ten years on. *Curr Opin Chem Biol* **15**: 489-496

Lee I and Schindelin H (2008) Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* **134**: 268-278

Lepre CA (2011) Practical aspects of NMR-based fragment screening. *Methods Enzymol* **493**: 219-239

Leslie AGW (2006) The integration of macromolecular diffraction data. *Acta Cryst* D**62**: 48-57

Li SJ and Hochstrasser M (1999) A new protease required for cell-cycle progression in yeast. *Nature* **398**: 246-251

Li S, Li Y, Carthew RW and Lai ZC (1997) Photoreceptor cell differentiation requires regulated proteolysis of the transcriptional receptor Tramtrack. *Cell* **90**: 469-478

Li S, Xu C and Carthew RW (2002) Phyllopod acts as an adaptor protein to link the sina ubiquitin ligase to the substrate protein tramtrack. *Mol Cell Biol* **22**: 6854-6865

Li T, Huang S, Dong M, Gui Y and Wu D (2012) Prognostic impact of SUMO-specific protease 1 (SENP1) in prostate cancer patients undergoing radical prostatectomy. *Urol Oncol* Epub ahead of print

Li W, Bengtson MH, Ulbrich A, Matsuda A, Reddy VA, Orth A, Chanda SK, Batalov S and Joazeiro CA (2008) Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signalling. *PloS One* **3**: e1487

Liang Y, Huimei Hong F, Ganesan P, Jiang S, Jauch R, Stanton LW and Kolatkar PR (2012) Structural analysis and dimerization profile of the SCAN domain of the pluripotency factor Zfp206. *Nucleic Acids Res* **40**: 8721-8732

Liani E, Eyal A, Avraham E, Shemer R, Szargel R, Berg D, Bornemann A, Riess O, Ross CA, Rott R and Engelender S (2004) Ubiquitylation of synphillin-1 and α-synuclein by SIAH and its presence in cellular inclusions and Lewy bodies imply a role in Parkinson's disease. *Proc Natl Acad Sci USA* **101**: 5500-5505

Lipinski CA, Lombardo F, Dominy BW and Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**: 3-26

Lipkowitz S and Weissman AM (2011) RINGs of good and evil: RING finger ubiquitin ligases at the crossroads of tumour suppression and oncogenesis. *Nat Rev Cancer* **11**: 629-643

Liu J, Stevens J, Rote CA, Yost HJ, Hu Y, Neufeld KL, White RL and Matsunami N (2001) Siah-1 mediates a novel β-catenin degradation pathway linking p53 to the adenomatous polyposis coli protein. *Mol Cell* **7**: 927-936

Liu M, Hsu J, Chan C, Li Z and Zhou Q (2012) The ubiquitin ligase Siah1 controls ELL2 stability and formation of super elongation complexes to modulate gene transcription. *Mol Cell* **46**: 325-334

Madu IG, Namanja AT, Su Y, Wong S, Li YJ and Chen Y (2013) Identification and characterization of a new chemotype of noncovalent SENP inhibitors. *ACS Chem Biol* Epub ahead of print

Mahajan R, Delphin C, Guan T, Gerace L and Melchior F (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**: 97-107

Mangani S, Carloni P and Orioli P (1992) Crystal structure of the complex between carboxypeptidase A and the biproduct analog inhibitor L-benzylsuccinate at 2.0 Å resolution. *J Mol Biol* **223**: 573-578

Marine JC and Lozano G (2010) Mdm2-mediated ubiquitylation: p53 and beyond. *Cell Death Differ* **17**: 93-102

Martens JH and Stunnenberg HG (2010) The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett* **584**: 2662-2669

Matsuzawa SI and Reed JC (2001) Siha-1, SIP, and Ebi collaborate in a novel pathway for β-catenin degradation linked to p53 responses. *Mol Cell* **7**: 915-926

Matsuzawa S, Takayama S, Froesch BA, Zapata JM and Reed JC (1998) p53-inducible human homologue of *Drosophila* seven in absentia (Siah) inhibits cell growth: suppression by BAG-1. *EMBO J* **17**: 2736-2747

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**: 491-497

Matunis MJ, Coutavas E and Blobel G (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J Cell Biol* **135:** 1457-1470

Mayer M and Meyer B (1999) A fast and sensitive method to characterize ligand binding by saturation transfer difference NMR spectra. *Angew Chem Int Ed Engl* **38**: 1784-1788

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* **40**: 658-674

Medhioub M, Vaury C, Hamelin R and Thomas G (2000) Lack of somatic mutation in the coding sequence of SIAH1in tumors hemizygous for this candidate tumor suppressor gene. *Int J Cancer* **87**: 794-797

Mohty B, El-Cheikh J, Yakoub-Agha I, Avet-Loiseau H, Moreau P and Mohty M (2012) Treatment strategies in relapsed and refractory multiple myeloma: a focus on drug sequencing and 'retreatment' approaches in the era of novel targets. *Leukemia* **26**: 73-85

Moldovan GL and D'Andrea AD (2009) FANCD2 hurdles the DNA interstrand crosslink. *Annu Rev Genet* **43**: 223-249

Möller A, House CM, Wong CS, Scanlon DB, Liu MC, Ronai Z and Bowtell DD (2009) Inhibition of Siah ubiquitin ligase function. *Oncogene* **28**: 289-296

Mossessova E and Lima CD (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* **5**: 865-876

Müller S, Hoege C, Pyrowolakis G and Jentsch S (2001) SUMO, ubiquitin's mysterious cousin. *Nat Rev Mol Cell Biol* **2**: 202-210

Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Winn MD, Long F and Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal strcutures. *Acta Cryst* D**67**: 355-367

Nadeau RJ, Toher JL, Yang X, Kovalenko D and Friesel R (2007) Regulation of Sprouty2 stability by mammalian Seven-in-Absentia homolog 2. *J Cell Biochem* **100**: 151-160

Nakayama K, Frew IJ, Hagensen M, Skals M, Habelhah H, Bhoumik A, Kadoya T, Erdjument-Bromage H, Tempst P, Frappell PB, Bowtell DD and Ronai Z (2004) Siah2 regulates stability of prolyl-hydroxylases, controls HIF1α abundance, and modulates physiological responses to hypoxia. *Cell* **117**: 941-952

Niesen FH, Berglund H and Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**: 2212-2221

Nissink JW (2009) Simple size-independent measure of ligand efficiency. *J Chem Inf Model* **49**: 1617-1622

Nordström H, Gossas T, Hämäläinen M, Källblad P, Nyström S, Wallberg H and Danielson UH (2008) Identification of MMP-12 inhibitors by using biosensor-based screening of a fragment library. *J Med Chem* **51**: 3449-3459

Okabe H, Satoh S, Furukawa Y, Kato T, Hasegawa S, Nakajima Y, Yamaoka Y and Nakamura Y (2003) Involvement of PEG10 in human hepatocellular carcinogenesis through interaction with SIAH1. *Cancer Res* **63**: 3043-3048

Orford K, Crockett C, Jensen JP, Weissman AM and Byers SW (1997) Serine phosphorylation-regulated ubiquitination and degradation of β-catenin. *J Biol Chem* **272**: 24735-24738

Otero A, Rodriguez de la Vega M, Tanco S, Lorenzo J, Aviles FX and Reverter D (2012) The novel structure of ac cytosolic M14 metallocarboxypeptidase (CCP) from *Pseudomonas aeruginosa*: a model for mammalian CCPs. *FASEB J* **26**: 3754-3764

Painter J and Merritt EA (2006) TLSMD web server for the generation of multi-group TLS models. J Appl Cryst **39**: 109-111

Park TJ, Hamanaka H, Ohshima T, Watanabe N, Mikoshiba K and Nukina N (2003) Inhibition of ubiquitin ligase Siah-1A by disabled-1. *Biochem Biophys Res Commun* **302**: 671-678

Passmore LA and Barford D (2004) Getting into position: the catalytic mechanisms of protein ubiquitylation. *Biochem J* **379**: 513-525

Peterson FC, Hayes PL, Waltner JK, Heisner AK, Jensen DR, Sander TL and Volkman BF (2006)Structure of the SCAN domain from the tumor suppressor protein MZF1. *J Mol Biol* **363**: 137-147

Polekhina G, House CM, Traficante N, Mackay JP, Relaix F, Sassoon DA, Parker MW and Bowtell DD (2002) Siah ubiquitin ligase is structurally related to TRAF and modulates TNF-α signalling. *Nat Struct Biol* **9**: 68-75

Praefcke GJ, Hofmann K and Dohmen RJ (2012) SUMO playing tag with ubiquitin. *Trends Biochem Sci* **37**: 23-31

Qi J, Kim H, Scortegagna M and Ronai ZA (2013) Regulators and effectors of siah ubiquitin ligases. *Cell Biochem Biophys* Epub ahead of print

Qi J, Pellecchia M and Ronai ZA (2010) The Siah2-HIF-FoxA2 axis in prostate cancer – new markers and therapeutic opportunities. *Oncotarget* **1**: 379-385

Rawlings ND, Barrett AJ and Bateman A (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **40**: D343-D350

Rees DC and Lipscomb WN (1981) Binding of ligands to the active site of carboxypeptidase A. *Proc Natl Acad Sci USA* **78**: 5455-5459

Relaix F, Wei XJ, Wu X and Sassoon DA (1998) Peg3/Pw1 is an imprinted gene involved in the TNF-NFκB signal transduction pathway. *Nat Genet* **18**: 287-291

Relaix F, Wei Xj, Li W, Pan J, Lin Y, Bowtell DD, Sassoon DA and Wu X (2000) Pw1/Peg3 is a potential cell death mediator and cooperates with Siah1a in p53-mediated apoptosis. *Proc Natl Acad Sci USA* **97**: 2105-2110

Relaix F, Weng X, Marazzi G, Yang E, Copeland N, Jenkins N, Spence SE and Sassoon D (1996) Pw1, a novel zinc finger gene implicated in the myogenic and neuronal lineages. *Dev Biol* **177**: 383-396

Reverter D and Lima CD (2004) A basis for SUMO protease specificity provided by analysis of human Senp2 and a Senp2-SUMO complex. *Structure* **12**: 1519-1531

Rodriguez de la Vega M, Sevilla RG, Hermoso A, Lorenzo J, Tanco S, Diez A, Fricker LD, Bautista JM and Aviles FX (2007) Nna1-like proteins are active metallocarboxypeptidases of a new and diverse M14 subfamily. *FASEB J* **21**: 851-865

Rodriguez MS, Dargemont C and Hay RT (2001) SUMO-1 conjugation *in vivo* requires both a consensus modification motif and nuclear targeting. *J Biol Chem* **276**: 12654-12659

Rogowski K, van Dijk J, Magiera MM, Bosc C, Deloulme JC, Bosson A, Peris L, Gold ND, Lacroix B, Grau MB, Bec N, Larroque C, Desagher S, Holzer M, Andrieux A, Moutin MJ and Janke C (2010) A family of protein-deglutamylating enzymes associated with neurodegeneration. *Cell* **143**: 564-578

Rotin D and Kumar S (2009) Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* **10**: 398-409

Roughley SD and Hubbard RE (2011) How well can fragments explore accessed chemical space? A case study from heat shock protein 90. *J Med Chem* **54**: 3989-4005

Rüdiger M, Plessman U, Klöppel KD, Wehland J and Weber K (1992) Class II tubulin, the major brain β tubulin isotype is polyglutamylated on glutamic acid residue 435. *FEBS Lett* **308**: 101-105

Rupp B (2010) Biomolecular crystallography: principles, practice, and application to structural biology ISBN 978-0-8153-4081-2

Sadowski M and Sarcevic B (2010) Mechanisms of mono- and poly-ubiquitination: Ubiquitination specificity depends on compatibility between the E2 catalytic core and amino acid residues proximal to the lysine. *Cell Div* **5**: 19

Sahab ZJ, Hall MD, Me Sung Y, Dakshanamurthy S, Ji Y, Kumar D and Byers SW (2010) Tumor suppressor RARRES1 interacts with cytoplasmic carboxypeptidase AGBL2 to regulate the α-tubulin tyrosination cycle. *Cancer Res* **71**: 1219-1228

Sainsbury S, Bird L, Rao V, Shepherd SM, Stuart DI, Hunter WN, Owens RJ and Ren J (2011) Crystal structures of penicillin-binding protein 3 from *Pseudomonas aeruginosa*: comparison of native and antibiotic-bound forms. *J Mol Biol* **405**: 173-184

Sander TL, Haas AL, Peterson MJ and Morris JF (2000) Identification of a novel SCAN box-related protein that interacts with MZF1B. The leucine-rich SCAN box mediates hetero- and homoprotein associations. *J Biol Chem* **275**: 12857-12867

Santelli E, Leone M, Li C, Fukushima T, Preece NE, Olson AJ, Ely KR, Reed JC, Pellecchia M, Liddington RC and Matsuzawa S (2005) Structural analysis of Siah-Siah-interacting protein interactions and insights into the assembly of an E3 ligase multiprotein complex. *J Biol Chem* **280:** 34278-34287

Scheffner M, Nuber U and Huibregtse JM (1995) Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**: 81-83

Schmidt RL, Park CH, Ahmed AU, Gundelach JH, Reed NR, Cheng S, Knudsen BE and Tang AH (2007) Inhibition of RAS-mediated transformation and tumourigenesis by targeting the downstream E3 ubiquitin ligase seven in absentia homolog. *Cancer Res* **67**: 11798-11810

Schneider G and Filipek A (2011) S100A6 binding protein and Siah-1 interacting protein (CacyBP/SIP): spotlight on properties and cellular function. *Amino Acids* **41**: 773-780

Schrödinger LCC (2010) The PyMOL Molecular Graphics System Version 1.5.0.4

Schumacher C, Wang H, Honer C, Ding W, Koehn J, Lawrence Q, Coulis CM, Wang LL, Ballinger D, Bowen BR and Wagner S (2000) The SCAN domain mediates selective oligomerisation. *J Biol Chem* **275**: 17173-17179

Scott DE, Coyne AG, Hudson SA and Abell C (2012) Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **51**: 4990-5003

Seeler JS and Dejean A (2003) Nuclear and unclear functions of SUMO. *Nat Rev Mol Cell Biol* **4**: 690-699

Seol JH, Feldman RM, Zachariae W, Shevchenko A, Correll CC, Lyapina S, Chi Y, Galova M, Claypool J, Sandmeyer S, Nasmyth K, Deshaies RJ, Shevchenko A and Deshaies RJ (1999) Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. *Genes Dev* **13**: 1614-1626

Shah JJ and Orlowski RZ (2009) Proteasome inhibitors in the treatment of multiple myeloma. *Leukemia* **23**: 1964-1979

Shah M, Stebbins JL, Dewing A, Qi J, Pellecchia M and Ronai ZA (2009) Inhibition of Siah2 ubiquiti ligase by vitamin K3 (menadione) attenuates hypoxia and MAPK signalling and blocks melanoma tumourigenesis. *Pigment Cell Melanoma Res* **22**: 799-808

Shen LN, Dong C, Liu H, Naismith JH and Hay RT (2006) The structure of SENP1-SUMO-2 complex suggests a structural basis for discrimination between SUMO paralogues during processing. *Biochem J* **397**: 279-288

Shen LN, Geoffroy MC, Jaffray EG and Hay RT (2009) Characterization of SENP7, a SUMO-2/3-specific isopeptidase. *Biochem J* **421**: 223-230

Shuker SB, Hajduk PJ, Meadows RP and Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**: 1531-1534

Siegel MG and Vieth M (2007) Drugs in other drugs: a new look at drugs as fragments. *Drug Discov Today* **12**: 71-79

Skowyra D, Craig KL, Tyers M, Elledge SJ and Harper JW (1997) F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**: 209-219

Sommer S, Weikart ND, Linne U and Mootz HD (2013) Covalent inhibition of SUMO and ubiquitin-specific cysteine proteases by an in situ thiol-alkyne addition. *Bioorg Med Chem* **21**: 2511-2517

Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**: 43R-56R

Susini L, Passer BJ, Amzallag-Elbaz N, Juven-Gershon T, Prieur S, Privat N, Tuynder M, Gendron MC, Israël A, Amson R, Oren M and Telerman A (2001) Siah-1 binds and regulates the function of Numb. *Proc Natl Acad Sci USA* **98**: 15067-15072

Tang AH, Neufeld TP, Kwan E and Rubin GM (1997) PHYL acts to down-regulate TTK88, a transcriptional repressor of neuronal cell fates, by a SINA-dependent mechanism. *Cell* **90:** 459-467

Tanikawa J, Ichikawa-Iwata E, Kanei-Ishii C and Ishii S (2001) Regulation of c-Myb activity by tumor suppressor p53. *Blood Cells Mol Dis* **27**: 479-482

The Maybridge Ro3 Fragment Library (www.maybridge.com)

Thiaville MM, Huang JM, Kim H, Ekram MB, Roh TY and Kim J (2013) DNA-binding motif and target genes of the imprinted transcription factor PEG3. *Gene* **512**: 314-320

Tounge BA and Parker MH (2011) Designing a diverse high-quality library for crystallography-based FBDD screening. *Methods Enzymol* **493**: 3-20

Ubersax JA and Ferrell JE Jr (2007) Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* **8**: 530-541

Ulrich HD (2009) The SUMO system: an overview. *Methods Mol Biol* **497**: 3-16

Uno M, Koma Y, Ban HS and Nakamura H (2012) Discovery of 1-[4-(N-benzylamino)phenyl]-3-phenylurea derivatives as non-peptidic selective SUMO-sentrin specific protease (SENP)1 inhibitors. *Bioorg Med Chem Lett* **22**: 5169-5173

Venables JP, Dalgliesh C, Paronetto MP, Skitt L, Thornton JK, Saunders PT, Sette C, Jones KT and Elliott DJ (2004) SIAH1 targets the alternative splicing factor T-STAR for degradation by the proteasome. *Hum Mol Genet* **13**: 1525-1534

Vendrell J, Aviles FX and Fricker LD (2004) (In) Messerschmidt A, Bode W and Cygler M (ed): Handbook of Metallopeptidases, John Wiley & Sons. pp. 167-192

Verger A, Perdomo J and Crossley M (2003) Modification with SUMO. A role in transcriptional regulation. *EMBO Rep* **4**: 137-142

Waby JS, Bingle CD and Corfe BM (2008) Post-translational control of sp-family transcription factors. *Curr Genomics* **9**: 301-311

Waites CL, Leal-Ortiz SA, Okerlund N, Dalke H, Fejtova A, Altrock WD, Gundelfinger ED and Garner CC (2013) Bassoon and Piccolo maintain synapse integrity by regulating protein ubiquitination and degradation. *EMBO J* **32**: 954-969

Walsh CT, Garneau-Tsodikova S and Gatto GJ Jr (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**: 7342-7372

Weber PC (1991) Physical principles of protein crystallization. *Adv Protein Chem* **41**: 1-36

Welcsh PL and King MC (2001) BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet* **10**: 705-713

Wells JA and McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**: 1001-1009

Wenlock MC, Austin RP, Barton P, Davis AM and Leeson PD (2003) A comparison of physiochemical property profiles of development and marketed oral drugs. *J Med Chem* **46**: 1250-1256

Wheeler TC, Chin LS, Li Y, Roudabush FL and Li L (2002) Regulation of Synaptophysin degradation be mammalian homologues of seven in absentia. *J Biol Chem* **277**: 10273-10282

Williams AJ, Blacklow SC and Collins T (1999) The zinc finger-associated SCAN box is a conserved oligomerisation domain. *Mol Cell Biol* **19**: 8526-8535

Williams AJ, Khachigian LM, Shows T and Collins T (1995) Isolation and characterization of a novel zinc-finger protein with transcription repressor activity. *J Biol Chem* **270**: 22143-22152

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A and Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Cryst* D**67**: 235-242

Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C, Sanschagrin F, Thomson NR, Winsor GL, Quail MA, Lennard N, Bignell A, Clarke L, Seeger K, Saunders D, Harris D, Parkhill J, Hancock RE, Brinkman FS and Levesque RC (2009) Newly introduced genomic prophage islands are critical determinants of *in vivo* competitiveness in the Liverpoo Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Res* **19**: 12-23

Winter M, Sombroek D, Dauth I, Moehlenbrink J, Scheuermann K, Crone J and Hofmann TG (2008) Control of HIPK2 stability by ubiquitin ligase Siah-1 and checkpoint kinases ATM and ATR. *Nat Cell Biol* **10**: 812-824

Wong CS and Möller A (2013) Siah: a promising anticancer target. *Cancer Res* **73**: 2400-2406

Wu H, Lin Y, Shi Y, Qian W, Tian Z, Yu Y and Huo K (2010) SIAH-1 interacts with mammalian polyhomeotic homologues HPH2 and affects its stability via the ubiquitin-proteasome pathway. *Biochem Biophys Res Commun* **397**: 391-396

Wu H, Shi Y, Lin Y, Qian W, Yu Y and Huo K (2011) Eukaryotic translation elongation factor 1 delta inhibits the ubiquitin ligase activity of SIAH-1. *Mol Cell Biochem* **357**: 209-215

Wu J, Tan X, Peng X, Yuan J and Qiang B (2003) Translocation and phosphorylation of calcyclin binding protein during retinoic acid-induced neuronal differentiation of neuroblastoma SH-SY5Y cells. *J Biochem Mol Biol* **36**: 354-358

Xu Z and Au SW (2005) Mapping residues of SUMO precursors essential in differential maturation by SUMO-specific protease, SENP1. *Biochem J* **386**: 325-330

Xu Z, Sproul A, Wang W, Kukekov N and Greene LA (2006) Siah1 interacts with the scaffold protein POSH to promote JNK activation and apoptosis. *J Biol Chem* **281**: 303-312

Yeh ET (2009) SUMOylation and De-SUMOylation: wrestling with life's processes. *J Biol Chem* **284**: 8223-8227

Yeh ET, Gong L and Kamitani T (2000) Ubiquitin-like proteins: new wines in new bottles. *Gene* **248**: 1-14

Yokoyama T and Nakamura T (2011) Tribbles in disease: Signaling pathways important for cellular function and neoplastic transformation. *Cancer Sci* **102**: 1115-1122

Yoshibayashi H, Okabe H, Satoh S, Hida K, Kawashima K, Hamasu S, Nomura A, Hasegawa S, Ikai I and Sakai Y (2007) SIAH1 causes growth arrest and apoptosis in hepatoma cells through β-catenin degradation-dependent and –independent mechanisms. *Oncol Rep*. **17**: 549-556

Yu HB, Kunarso G, Hong FH and Stanton LW (2009) Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. *J Biol Chem* **284**: 31327-31335

Yun S, Möller A, Chae SK, Hong WP, Bae YJ, Bowtell DD, Ryu SH and Suh PG (2008) Siah proteins induce the epidermal growth factor-dependent degradation of phospholipase Cε. *J Biol Chem* **283**: 1034-1042

Zhang J, Guenther MG, Carthew RW and Lazar MA (1998) Proteasomal regulation of nuclear receptor corepressor-mediated repression. *Genes Dev* **12**: 1775-1780

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40

Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepp DM, Elledge SJ, Pagano M, Conaway RC, Conaway JW, Harper JW and Pavletich NP (2002) Structure of the Cul1-Rbx1-Skp1-F box Skp2 SCF ubiquitin ligase complex. *Nature* **416**: 703-709

Zhou Y, Li L, Liu Q, Xing G, Kuai X, Sun J, Yin X, Wang J, Zhang L and He F (2008) E3 ubiquitin ligase SIAH1 mediates ubiquitination and degradation of TRB3. *Cell Signal* **20**: 942-948

Appendix: list of publications

Rimsa V, Eadsforth TC and Hunter WN (2011) The role of Co^{2+} in the crystallization of human SENP1 and comments on the limitations of automated refinement protocols. *Acta Cryst* F**67**: 442-445

Rimsa V, Eadsforth TC and Hunter WN (2013) Structure of the SCAN domain of human paternally expressed gene 3 protein. *PLoS ONE* **8**: e69538

Rimsa V, Eadsforth TC, Joosten RP and Hunter WN (2014) High-resolution structure of the M14-type cytosolic carboxypeptidase from *Burkholderia cenocepacia* refined exploiting PDB_REDO strategies. *Acta Cryst* D**70**

Rimsa V, Eadsforth TC and Hunter WN (2014) Two high-resolution structures of the human E3 ubiquitin ligase Siah1. *Acta Cryst* F